

STIC-Biotech/ChemLib

71965

From: Helmer, Georgia  
Sent: Monday, July 29, 2002 10:59 AM  
To: STIC-Biotech/ChemLib  
Subject: 09/643,755

Could you please do sequence searches of SEQ ID 1, 2, and 3 of this case, 09/643,755.

SEQ ID 1 and 3 are DNA. SEQ ID 2 is protein.

Could you also do an oligo search of SEQ ID 1 and 3.

The commercial and inhouse databases.

Thanks in advance for you assistance.

*Cheymsen*

Georgia L. Helmer Ph.D.  
Patent Examiner  
Crystal Mall 1, 9D14  
AU 1638  
703-308-7023  
Georgia.Helmer@USPTO.gov  
mailbox 9e12

*SPRMS* 61.

Point of Contact:  
Toby Port  
Technical Info. Specialist  
CM1 6A04  
703-308-3534

03/11/03  
591

Searcher: \_\_\_\_\_  
Phone: \_\_\_\_\_  
Location: \_\_\_\_\_  
Date Picked Up: 7/31  
Date Completed: 8/2  
Searcher Prep/Review: 12  
Clerical: \_\_\_\_\_  
Online time: 12

TYPE OF SEARCH:  
NA Sequences: 4  
AA Sequences: 1  
Structures: \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Full text: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

VENDOR/COST (where applic.)  
STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
Questel/Orbit: \_\_\_\_\_  
DRLink: \_\_\_\_\_  
Lexis/Nexis: \_\_\_\_\_  
Sequence Sys.: 48  
WWW/Internet: \_\_\_\_\_  
Other (specify): \_\_\_\_\_

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 12:44:09 ; Search time 9237.8 Seconds  
(without alignments)  
2657.216 Million cell updates/sec

Title: US-09-643-755B-1

Perfect score: 1173

Sequence: 1 atgaattccttaagctctt.....gctagctaaagaatctga 1173

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenBank:\*

1: gb\_da:\*

2: gb\_hcg:\*

3: gb\_in:\*

4: gb\_om:\*

5: gb\_ov:\*

6: gb\_pat:\*

7: gb\_ph:\*

8: gb\_pl:\*

9: gb\_pr:\*

10: gb\_ro:\*

11: gb\_sts:\*

12: gb\_sy:\*

13: gb\_un:\*

14: gb\_vl:\*

15: em\_da:\*

16: em\_fun:\*

17: em\_hum:\*

18: em\_in:\*

19: em\_mu:\*

20: em\_om:\*

21: em\_or:\*

22: em\_ov:\*

23: em\_pat:\*

24: em\_ph:\*

25: em\_pl:\*

26: em\_ro:\*

27: em\_sts:\*

28: em\_un:\*

29: em\_vl:\*

30: em\_htg\_hum:\*

31: em\_htg\_inv:\*

32: em\_htg\_other:\*

33: em\_htgo\_inv:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Query	Score	Match Length	DB ID	Description
------------	-------	-------	--------------	-------	-------------

1	1173	100.0	1173	6	AX088019	AX088019 Sequence
2	1173	100.0	3957	6	AX088021	AX088021 Sequence
3	900.4	76.8	1305	4	BOYCHMOB	J00003 bovine chym
4	897.2	76.5	1240	6	AR002347	AR002347 Sequence
5	897.2	76.5	1460	6	E00042	E00042 DNA coding
6	896.2	76.4	1291	6	A15836	A15836 chymosin ge
7	895.6	76.4	1175	6	I04058	I04058 Sequence 5
8	894.8	76.3	2733	6	AR073077	AR073077 Sequence
9	894	76.2	1275	4	BOYCHMOA	J00002 bovine chym
10	894	76.2	1290	6	E00075	E00075 CDNA encodi
11	894	76.2	1291	6	A15633	A15633 preprochymo
12	893	76.1	1289	6	E00144	E00144 CDNA encodi
13	893	76.1	2726	6	I08097	I08097 Sequence 5
14	892.4	76.1	1311	6	E00108	E00108 DNA coding
15	892.4	76.1	1460	6	E00295	E00295 CDNA encodi
16	890.8	75.9	1269	4	BOYCHMOC	J00004 bovine chym
17	886.8	75.6	1098	6	E00079	E00079 CDNA sequen
18	886.8	75.6	1098	6	E02341	E02341 CDNA sequen
19	872.8	74.4	2982	6	I08098	I08098 Sequence 1
20	870.8	74.2	1101	4	AF177290	AF177290 Bubalus a
21	839.6	71.6	1292	4	OAPPCY	X53037 Lamb mRNA f
22	776	66.2	1115	12	SNPROCA	M22593 Synthetic b
23	753.6	64.2	1094	4	BP019786	U19786 Bos primige
24	749	63.9	1172	4	CDR131677	AJ131677 Camelus d
25	732.4	62.4	1230	4	SSU14406	U14406 Sus scrofa
26	718	61.2	1262	9	AB038386	AB038386 Callithr1
27	627.2	53.5	969	6	A02000	A02000 Artificial
28	627.2	53.5	969	6	A06446	A06446 Artificial
29	619.6	52.8	1146	12	SNPROCA	M21016 Synthetic b
30	605	51.6	1259	10	RNO251688	AJ251688 Rattus no
31	534	45.5	1140	6	AX19837	AX19837 Sequence
32	382.4	32.6	637	6	E05472	E05472 DNA sequence
33	381.8	32.5	1403	4	AB047245	AB047245 Canis fam
34	378.6	32.3	1161	4	AB047245	AB047245 Rhinoloph
35	373.8	31.9	1351	9	AB038384	AB038384 Callithr1
36	370.8	31.6	1381	4	AB047244	AB047244 Sorex ung
37	363.8	31.0	1362	4	AB047243	AB047243 Suncus mu
38	363.8	31.0	1363	4	PIGPEPA	J04601 Pig pepsiho
39	363.4	31.0	1221	4	CDR131678	AJ131678 Camelus d
40	362.2	30.9	1276	5	CHKPSN	D00215 Gallus gall
41	360.8	30.8	1266	4	RABPEP123	M59235 Rabbit peps
42	357.4	30.5	1353	9	MEPEPA23	X59755 M. fuscata m
43	357.4	30.5	1361	9	MEPEPA4	X59753 M. fuscata m
44	353.6	30.1	1371	4	RABPEP111	M59237 Rabbit peps
45	352.6	30.1	1338	9	MACPCA	M20788 Rhesus monk

## ALIGNMENTS

RESULT 1

AX088019

LOCUS AX088019

DEFINITION Sequence 1 from Patent WO0114571.

ACCESSION AX088019

VERSION AX088019.1

KEYWORDS GI:13396947

SOURCE

ORGANISM

Bos taurus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bos.

REFERENCE

1 (bases 1 to 1173)

AUTHORS van Rooijen, G., Keon, R. G., Boothe, J. and Shen, Y.

TITLE Commercial Production of Chymosin in Plants

JOURNAL Patent: WO 0114571-A 1 01-MAR-2001;

SEMIOSIS Genetics Inc. (CA)

FEATURES

source

1..1173

/organism="Bos taurus"

/db\_xref="taxon:9913"

1..1173

/note="unnamed protein product"

```

/codon_start=1
/protein_id="CAC34680.1"
/db_xref="GI:1336948"
/translation="MNFLLKSPFYAFLFCGOYFVAAYTHAAETIRIPLYKGLSKRLALK
EHGLLEDFLOKOOYGISKSYSGFEVAVSLPTNLNDISOPEKILYIGPPPODFYALFPTD
GSSDFWVSLICKSNACKNHQRFDPKRSFTFONLNGKPLSIHYGSGMOGLIGDTYIVY
SNVIDIOATVGLSTOEPGDVTFYAEFGIIGMAVPLASEYSIIVFNNMNRHLVAD
LFSYMDRNGSESMLTGGLIDIPSYTSGLHWPTFOWQOFTDVSYLSISGVVACGKG
GCGLILDTGTSKIVGSPSDILNIOALIGANOVQGEFDIDCNDISMTPEYVLEINMGK
YPLPSPAYTSDDOFCFTSGFSGEHSQKWLGDVIRIRIYISVFPRANNLGLARAT"
299 a 308 c 262 g 304 t

```

Query Match	100.0%	Score 1173	DB 6	Length 1173
Best Local Similarity	100.0%	Pred. No. 0		
Matches 1173, Conservative	0	Mismatches	0	Indels 0; Gaps 0

OY	1	atbaacttccttaagctcttcctcccttctaagctcttccttggttcggttaactacgtc	60
Db	1	ATGAACCTCCCTTAAGCTCTTCCTTTCTTACCGCTTCCTTTGTTCCGTCATTAATCTTCGT	60
OY	61	gcgttactacacgctgtcgtagatcacccgcacctctctctacaaaggtaaagtctctcgt	120
Db	61	GCTGTACTACCGCTGCCTGAGATCACCCGCATTCCTCTCTACAAAGGTAAAGTCTCTCCGT	120
OY	121	aagcgctggaagaaacatbgaacctctagaagactctctgcaagaacaacagtatgcac	180
Db	121	AAGCGCTGGAAGGAACATGAGCTCTTGAAAGACTCTCTTGAGAAACAAACAGTATGGCATC	180
OY	181	agcagcaagtaactcccgcttcggtgaagtctgtagcgctgcacattacaactacacacttgat	240
Db	181	AGCAGCAAGTACTCCGGCTTGCGGGAAGTTCTGTAGCGCTGCCACTATCCACTACCTTGAT	240
OY	241	agtcaataccttgggaaagatctacctcgaaaccccgctcaagaagttacacggtctctt	300
Db	241	AGTCATACCTTTGGGAAGATCTTACCTCGAACCCTCCGCTCAAGAGTTACACGTTCTCTTT	300
OY	301	gaiactggcttcctctgactcttcgttgcctctctatactacgcaagaaagtgcctgcaag	360
Db	301	GATACTGGTTCTCTTGACTTCTTGGGTTTCCCTCTATCTACTGCAAGAACAAATGCTTCGAAG	360
OY	361	aaccaccaagaatttcgaatccgagaagaatcgclccaactctccaagaacttaggcaaaccttg	420
Db	361	AACCAACAAAGATTCGATCCGAGAAAGTCGNCACCTTCCAGAACTTAGCCAAACCTTGG	420
OY	421	tctatacactacggtacaagttagatagaatgcagaagaatctctaggtctatgatataccgtcaactg	480
Db	421	TCTATACCTTCGGGTACAGGATGAGCATCCAAAGAACTCTTAGCATATGATACCGTACACTGTC	480
OY	481	tccaacacttgggacatccaacaacagtaggacttagcaaccaagaacaggttgatg	540
Db	481	TCCAACTTTGGGACATTCAACACACAGTAGACTTATGCAACCCAGAACACAGGTGATGTC	540
OY	541	ttcaactatgcaagaattcgatgtagctccttggtatggtcatataccatcgctcggtcaag	600
Db	541	TTCACTTATGCAAGATTCGATGAGGATCCTTGGTATGGCATATCCATCGCTCGGTCAGAG	600
OY	601	tactcgataacttggtttgtaaaaaatgatagaaccgaacacttagtagcttaagaactggtc	660
Db	601	TACTCGATACCTGTGTTTGACAAACATATGAAACCGACACCTTAGTACCTCAAGACTGTTTC	660
OY	661	tcggtttacaagcaagaatagcagagagaagcatgctcaacgcttggagatatgataca	720
Db	661	TCGGTTTACATGAGACAGGAATGGCCAGSAGAGCATGCTCACGCTTGGAGCTTATTTGATCCA	720
OY	721	tctactatacacagaatcctctcaacttgggttccagtgcaatgltgcaagcagttactggaattc	780
Db	721	TCCTACTACACAGAGATCTCTTCACTGGGTTCCAGTCACTGTCAGCAGCAGTACTGGCAATTC	780
OY	781	acgtgtgacagtgatccaccatcaagcggtgtggtgtgtgcagtgatgaaggtgatatcaagc	840
Db	781	ACTGTGACACTGTACCATTCATAGGGGTGTGGTTTGTTCATGATGAAGGTGATGTCAAGCT	840

Oy	841	atcttgataccggtgaagctccaagcttgctggacttagcagcgacatcttcaaatctag	900
Db	841	actcttgataccggtgactcccaagctggtcagcttaccacgacattctcaacttttag	900
Oy	901	caagctatctgagagccacacagacnaccagctgtagtggatcttgacatagattcgacaactt	960
Db	901	caagctattttggagccacacagacnaccagctgtagtggatcttgacatgattggcacactt	960
Oy	961	agctacatgctacacgttgcctcttgaatacaagcgcaagatgatacccactgacccctcc	1020
Db	961	agctacatgctcctacacgttgcctcttgaatacaagcgcaagatgatacccactgacccctcc	1020
Oy	1021	gacctataccagccagatccaaggtctctgacacagctggaattccagatgtagaacattcc	1080
Db	1021	gccctttaccagccagatccaaggtctctgacacagctggaattccagatgtagaacattcc	1080
Oy	1081	cagaaatggtatcttggggagatggtgtcatcttctgtgagtactacagctcttggacagggcc	1140
Db	1081	cagaaatggtatcttggggagatggtgtcatcttctgtgagtactacagctcttggacagggcc	1140
Oy	1141	aacaacctcgtttggctagctaaagcaatcttga	1173
Db	1141	aacaacctcgtttggctagctaaagcaatcttga	1173

	RESULT	2				
	AX088021					
	LOCUS		3957 bp	DNA	linear	PAT 17-MAR-2001
	DEFINITION	Sequence from Patent WO0114571.				
	ACCESSION	AX088021				
	VERSION	AX088021.1	GI:13396949			

SOURCE	ORGANISM
synthetic construct.	
synthetic construct.	
artificial sequence.	
1 (bases 1 to 5957).	
REFERENCE	
AUTHORS	van Rooijen, G., Keen, R.G., Boothe, J. and Shen, Y
TITLE	Commercial production of chymosin in plants
JOURNAL	Patent: WO 014571-A 3 01-MAR-2001.

**SemBioSys Genetics Inc. (CA)**  
Location/Qualifiers

```

/organism="synthetic construct
/db_xref="taxon:32630"
1554. .2726
CDS

```

```

/protein_id="CAC34681.1"
/db_xref="gi:13396950"
/translation="MNFLLSKSPGEFAFLCGFYEVAVTHAAETIRPIPLYKKSLRKALK
EHGLLEDFLOKOOXGISSKYSGFEVAVSLPTLNTLDOSGYKILGLPPDEFLPTPT
GSSFWVPTCKRNACKNRRDPRKSPFONIGRLSIHGXGSGMGLIGADTVAV
SNLIDIOQTGELSLQEPGDVPTVAEEFGILLKAVTSLASESTIVFPMNMRHLYAD
LFSVYMRNGEESMLTGALIDPSTITSLSHVAFTVQIMQFTVDSTISGVVYACGG
GCAQALITGTSKLVGPSDDLINQALGAIQNOQDEGDLDDNLSITVPTVFEINGM
YLPSPASVTSDDQSCFSGSGFSEHNSKILGDVETREEDYSIVFERRANLLAKAT"
BASE COUNT      1263 a      790 c      609 g      1295 t
BRGIN

```

Query Match	100.0%	Score 1173	DB 6	Length 3957
Best Local Similarity	100.0%	Pred. No. 0		
Matches 1173; Conservative	0	Mismatches	0	Indels 0; Gaps 0

QY	1	atgaacttccttaagtccttccctctcttaagccttccctctatttcggtacaaactcgtt	60
Db	1554	ATGAACCTTCCTTAAGTCCTTCCCTCTCTTAAGCCTTCCTCTATTTCGGTAAACTTCGTT	16133
QY	61	gccttactacagctgcttgatgatacccgatctctctctctaaagaagtatgctctcgt	120
Db	1614	GCTGTATATCATCGCTGCTGGAGATCACCGGATTCCTCTCTTAACAAGGTAAAGTCTCCGT	16733



QY 121 aagcgctgaaggaacatgagcttctagaagacttcttcgagaacaacagatgacac 180  
|||||  
Db 1674 AAGGGGCGGAAGGAACAATGCACTTCTAGAACACTTCTTGCAAGAAACAACACTATGCAATC 1733  
QY 181 agcagcaagtactccgcgcttcgtgtaagttgctagcgtgcacttaccactacattgat 240  
|||||  
Db 1734 AGCAGCAAGTACTCCGGCTTGCTGAAGTTGCTAGCCGTGACACTTACCACTACCTTGAT 1793  
QY 241 agtaataactttgggaagatctacctcgaaaccccgctcgaagattaccggttctctt 300  
|||||  
Db 1794 AGTAATAACTTTGGGAAGATCTACCTCGGAACCCCGCTCAAGAGTTCCACGGTTCTCTT 1853  
QY 301 gatactggttcctctgacttcgtggtttccctctatctactcaagaagaatgcctgcag 360  
|||||  
Db 1854 GATACGTGTTCTCTGACTTCTGGTTCCCTCTATCTACTCAAGACGAATGCTCGCAAG 1913  
QY 361 aaccacaagaatccgatccgagagaagtcgtccacctccgaacttaggcaaaccttg 420  
|||||  
Db 1914 AACCAACCAAGATTCGATCCGAGAAAGTCGTCCACCTTCCAGAACTTAGGCAAAACCCCTTG 1973  
QY 421 tctatacctacggttacaggttagcatgcaagaatcttagctatgataccgctacgttc 480  
|||||  
Db 1974 TCTTACACTACGCTACAGGTAGCATGCAAGGAATCTTAGGCTATGATACCGCTCACTGTC 2033  
QY 481 tccaacattgtgacatccaacagacagtagacttagcaccacaagaacaggtgattgc 540  
|||||  
Db 2034 TCCAAACATTTGGAGCATTCACAGACAGTAGACTTAGCACCAAGAACGAGTGATGTC 2093  
QY 541 ttcacatctgcagaatccgatgcacatccctgtgtatggtacataccatcgctgcagag 600  
|||||  
Db 2094 TTTCACCTATGAGAAATTCGATGCGATCCTTGATGCGATACCCATCCCTGCGCTCAAG 2153  
QY 601 taccgcgaatccgtgtttgacaacatgatagcagacacccatgagctcgaagctgttc 660  
|||||  
Db 2154 TACTCGATACCTGTGTTTGAACAATGATGAAACCGACACCTTAGTACCTCAAGACTGTC 2213  
QY 661 tccggttatacagcaaggaatgagccagagagacatgctcaacgcttggagactatgatca 720  
|||||  
Db 2214 TCGGTTTACATGAGCAAGGAATGGCCAGAGAGCATGCTCAAGCTTATGATGCA 2273  
QY 721 tccactacaagaagatctctcaactggttccagtcacgtgtgcagcagtagctggcaattc 780  
|||||  
Db 2274 TCCACTACACAGAGATCTCTTCACTGCGGTTCCACTGCTGACAGCACTACTGCAATTC 2333  
QY 781 actgtgacaggtgcacacatcagcgtgtgtgttgcattgtagagtgagttgcagct 840  
|||||  
Db 2334 ACTGTGACACTGTCAACCATCAGCGGTGTGTGTGATGTGAAGGTGGAATGTCAGCT 2393  
QY 841 atcttgataccggttaagtcacagctgtgcagactagcagcagacatctcaaatcag 900  
|||||  
Db 2394 ATCTTGGATACCGGTAGCTCAAGCTGCTGCGACCTAGCAAGCATTTCTCAATTCAG 2453  
QY 901 caagctattggagccacacagaaacagtaecgtgtagattgacatagattgcgaacctt 960  
|||||  
Db 2454 CAACCTTGTGGAGCAACAAGAACACAGTACGCTGAGTTTGACATAGATTGCGACAACCTT 2513  
QY 961 agctacatgctctacagttgtctttgagatcaaggaagatgtaaccactgacccccctcc 1020  
|||||  
Db 2514 AGCTACATGCTTACAGTTGTCTTTGAGATCAACGGCAAGATGTACCCACTGACCCCTCC 2573  
QY 1021 gctataccagccagatcaagggttctgcaacagtgatctcagagatggaacattcc 1080  
|||||  
Db 2574 GCTATACACAGCCAGATCAAGGCTTCTGCACAGTGGAATTCCAGAGTGAACAACATTC 2633  
QY 1081 caggaataggtatctggagagatggttcaattcgttgagtactacagcgtctcttgacaggcc 1140  
|||||  
Db 2634 CAGAAATGGATCTTGGAGAGATGTGTATTTGTGATGATCTACAGCGTCTTGGACAGGGCC 2693  
QY 1141 aacaacctcgttgggtagactaaagcaatctga 1173  
|||||  
Db 2694 AACCACTCTGTTGGGCTTAGCTAAAGCAATCTGA 2726

RESULT 3  
BOVCHYMOB  
LOCUS BOVCHYMOB 1305 bp mRNA linear MAM 26-APR-1993  
DEFINITION bovine chymosin b (rennin) mrna.  
ACCESSION J00003  
VERSION J00003.1 GI:162859  
KEYWORDS chymosin; chymosin B; protease; rennin.  
SOURCE bovine (calf) cdna of fourth stomach mucosa mrna.  
ORGANISM Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;  
Bovidae; Bovinae; Bos.  
REFERENCE  
AUTHORS Harris,T.J.R., Lowe,P.A., Lyons,A., Thomas,P.G., Eaton,M.A.W.,  
Mulligan,T.A., Patel,T.P., Bose,C.C., Carey,N.H., and Doel,M.T.,  
TITLE Molecular cloning and nucleotide sequence of cdna coding for calf  
preprochymosin  
JOURNAL Nucleic Acids Res. 10, 2177-2187 (1982)  
MEDLINE 82221400  
COMMENT chymosin is the major proteolytic enzyme in the fourth stomach of  
the unweaned calf. two chromatographically different forms, a and  
b, of the enzyme and its precursor are known and a third form seems  
likely (see bovchymoa, bovchymoc). this sequence has tentatively  
been identified as coding for preprochymosin b. sequence comparison  
indicates that the precursors for chymosins a and b differ by only  
two amino acids, and for b and c by only four amino acids.  
FEATURES  
source  
1..1305  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
<1..1305  
/product="chymob mRNA"  
26..1171  
/note="preprochymosin b"  
/codon\_start=1  
/protein\_id="AAA30448.1"  
/db\_xref="GI:162860"  
/translation="MRCLVYLAVFLASGAEITRTPLYKSKLRKRGHLEDT  
OKDOXGSSKSYSGFEVAPLPIINLDSQFGRKTYLCTPPQETVFLDTSSSPWPS  
ITKSNACKNHGRFDPKRSSTFONLGRPLSIHGTSQMGILGIVTVSNIVDIQT  
VGLSTQEPGVFVAEFDGILGMAVPSLAEYSIPVDNMMNRLVAQDLFSYMDRN  
GOESMLTGLAINPSYTGSLHWPVYCOQTVDSVTSISGVVACEGSCQAILDTG  
TSKIVGPSDLINQOALIGATONQGFQIDCNLSYMPVIVEINKMPLPMSAVT  
SDQGFCTSGFQSEHNSQKMLIDVFIREYYSVFDNRNNLVGLAKAI"  
29..773  
/note="presequence"  
mat\_peptide  
/product="chymosin b"  
BASE COUNT 305 a 393 c 340 g 267 t  
ORIGIN  
pst-i site.  
Query Match 76.8%; Score 900.4; DB 4; Length 1305;  
Best Local Similarity 87.8%; Pred. No. 5e-263;  
Matches 982; Conservative 0; Mismatches 136; Indels 0; Gaps 0;  
56 tcgtgtcgttactccaogctgtcgtatgataccgcacatctccttacaagaagtaagctc 115  
|||||  
Db 54 TCTTGGCTCTCTCCCAAGGCGCTGAGATCACACAGATCCCTCTTACAAAGCAAGTCTC 113  
QY 116 tccgtaagcgctgaaggaacatgactctagaagactcttcagagaacaacagatg 175  
|||||  
Db 114 TGAGGAAGGCCCTTAAGAGAGATGGCTTCTGGAGACTTCTCTCAGAAACAGCAGTATG 173  
QY 176 gcatcagcagcaagtaactccggtcttcgtgaagttgctagcgttgcacttaaccaactacc 235  
|||||  
Db 174 GCATCAGCAGCAAGTACTCCGGCTTGCGGAGGTGGCCAGCGGTGCCCTGTACCAATGACC 233  
QY 236 ttgtagtcaatactttgggaagaatctaccctcggaaccccgctccaagaagtaccgctc 295  
|||||  
Db 234 TGGATAGTCACTATTGTTGGGAAGATCTACCTCGGAGACCCGCCCCAGAGATTCAACCGTC 293

Qy	296	tccttgataacggtgtccctctgaactcttggtgtccctcctatctactgtgaaagcgaaatgct	355
Db	294	tggtttgacacagtgccctctgactgttctggtaacctctatctactgcgaagacaaatgct	353
Qy	356	gcaagaacccacaaagatctgcatccgagaagtctgtccacctccagacttaagcaaac	415
Db	354	gcataaaacctacagcgtttcgaccocgagaaagatgctgacactctccagaaacctggcgaac	413
Qy	416	ccctgtctatacactaagtgatcaggtagcatatgcgaagaaattcttggctatgatcatctca	475
Db	414	ccctgtctatctacactacggaagcagacacacagcagggcattctgagcgtatgacacccgtca	473
Qy	476	ctgtctccaacatctgtgacatctcaacagacagtagagacttagaaccccaagaacacagtg	535
Db	474	ctgtctccacacatctgttgacatctccagcagacagtagagccttagacacccaggaagccgggg	533
Qy	536	atgtctcaacatctagcagaattcgalatgcctccttgytatagcatlaccatcgcctgcgt	595
Db	534	acgcttccacactatgcccgaatttcgacggggaatctgggggattggccttaccctctgcctgcct	593
Qy	596	cagagtagctgcatacctctgttttgaacaacatgataagaacccagacacttagtaactaaagct	655
Db	594	cagatctactgcatacctcctgttttgacaacatgatagaacagcacccttggtggcccaagacc	653
Qy	656	tgcttcctggtttacatgagcaggaatggccaggaagacatgctcaagcgtcttgaagatctg	715
Db	654	tggtttcctggtttacatgagcaggaatggccaggaagacatgctcaagcgtctgagggggcatca	713
Qy	716	atccatctactacaagaagatctctctactctggtttccacgtacactgtagagcaagtactggc	775
Db	714	accctctctactacacagaagtgctccctgcacactgggtgcccgttaacatgtagagcaagtactggc	773
Qy	776	aattcacctgtagacagtgctcacacatcaagcggctgtgtgttctgtcatgtgaagtgatgtc	835
Db	774	agttcacctgtagacagtgctcacacatcaagcgggtgtgtgtgtgtgtgaggggtggcgtgtc	833
Qy	836	aagatctcttgtagataccggtatcgtccaaagctggtctgcgaacttagcagagcaattctcaaca	895
Db	834	aggccattcttgtagataccggtatcgtccaaagctggtctgcgaacttagcagagcaattctcaaca	893
Qy	896	ttcacgaagctatctgagacacacagaacacagtaacgtgaatgtttgacatagatctgcga	955
Db	894	tgcacgaagctatctgagacacacagaacacagtaacgtgaatgtttgacatagatctgcga	953
Qy	956	accctagctacaatgcctacagttgtctcttgtagatcaacgcgcaagaatgtlaccactgaccc	1015
Db	954	acctagctacaatgcctacagttgtctcttgtagatcaacgcgcaagaatgtlaccactgaccc	1013
Qy	1016	cctgcgcctataccgacgacgaatacaaggtgtctgcacacagctgagattccagaagtgtagaac	1075
Db	1014	cctgcgcctataccgacgacgaatacaaggtgtctgcacacagctgagattccagaagtgtagaac	1073
Qy	1076	attcccaagaatagatctcttggaagatgtgttctcatctcgtgaatactaactacgctctttgaca	1135
Db	1074	attcccaagaatagatctcttggaagatgtgttctcatctcgtgaatactaactacgctctttgaca	1133
Qy	1136	ggggccaacaacactcgtgttggtcgtactgaagaacatctga	1173
Db	1134	ggggccaacaacactcgtgttggtcgtactgaagaacatctga	1171
RESULT 4			
LOCUS	AR002347	1240 bp	DNA
DEFINITION	Sequence 2 from patent US 574165.	linear	PAT 04-DEC-1998
ACCESSION	AR002347		
VERSION	AR002347.1	GI:3963901	
KEYWORDS			
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	1 (bases 1 to 1240)		
AUTHORS	Kato, E.K. and Stuart, W. Dorsey.		

TITLE	Light-regulated promoters for production of heterologous proteins
JOURNAL	in filamentous fungi
Patent:	US 5741665-A 2 21-APR-1998;
FEATURES	Location/Qualifiers
source	1. .1240
BASE COUNT	/organism="unknown"
ORIGIN	274 a 374 c 339 g 253 t
Query Match	76.5%; Score 897.2; DB 6; Length 1240;
Best Local Similarity	87.7%; Pred. No. 4.7e-262;
Matches 980; Conservative	0; Mismatches 138; Indels 0; Gaps 0;
OY	56 tctgtgcgttactacagctcgtctgtgatccaccgattccctctcttaacaagttaagcttc 115
Db	100 tctttgcctcttccccagggccttgagttaccagatccctcttgcacaaaggcattctc 159
OY	116 tccgttaagcgctgtaagaagacatgtaactcttgaagaactctcttcagaaacacagatg 175
Db	160 tgaggaagggccctgtaaggagcatggcgctttcttgagagactccttcgagaAACGACGATG 219
OY	176 gcatcaagcagcaagtaactcccgctctcgttgtaagttgtctagogtgcacttaccactacc 235
Db	220 gcattAGCAGCAAGTACTCCGGCTCGGGGAGTGGCCAGCGCTGCCCTGACCAACTAC 279
OY	236 ttgatatgaataactcttggaaagatctaaccttcgaaaccccgctctaagaagttcacgcttc 295
Db	280 tggatagttcagtaacttttgggaagatctactctgggaaccccgccgccagagattcACGTTG 339
OY	296 tctttgatactgtgtctctctgactcttcgttggttcctctatactatgcaagaaatgcct 355
Db	340 tctttgacacttgccctctctgtactcttctgggtaccctctatctactgcaAAGCAATGCT 399
OY	356 gcaagaacacccaagaatttcgaatcgaagaagttgtccacttcagaacttaggcaaac 415
Db	400 gcaaaAACCCACGAGCGCTTCGACCcGAGAAAGTcGTCCACTTCCAGAACTCGGGCAAGC 459
OY	416 ccttgcctatacactcagtaagtlagcatgcaagaactcttaagctatgatatccgtca 475
Db	460 cccctgtctATCCACtACAGGGAGCAGGAGCATGCAAGGATCCTTAGGCTATGACACCGTCA 519
OY	476 cgtgttccaaacttttgaaacttaaaagacagatgagacttgaaccccaagaacagctg 535
Db	520 cgtgtccCAACTTGTGGATCCACGAGACAGtAGGcCTTGAGCACCAGGAGCCCGGG 579
OY	536 atgtctcaaccatgacgaattcgaatgycatcccttgatagtcgacatacccatgcgtcgct 595
Db	580 acgtgttcACtATGCGCGAAATTGcAGCGGgATCTTGGGGATGGCTTACCCCTCGTCGCT 639
OY	596 cagaagtactcgataacctgtgtttgacaacatgataaaccgaacacttagtaagctaaagct 655
Db	640 cAGAGTACTCGATACCCCGTtTGACAAATATATAAACGGGACCTGGTGCGCCAAAGAC 699
OY	656 tttcttcggttttaacttgagcagaatggccagagagacatgctcaacgctttgagctatg 715
Db	700 tcttctcGgTTTACATGAGCAGAGAAATGGCCAGAGACATGCTCACGCTGGGGGCCATMG 759
OY	716 atccatctactatacaagaatctcttcacatggtttccagtaactcgttgcaagcagtaatggc 775
Db	760 ACCCGTCTACTACACAGAGGTCCTTGCACTGGGTGCCCTCGTACAGTGGACGATGCTGC 819
OY	776 aattcaactgttgaaacagttgcaccatcaagggtgtgtgtttgttcagtlgaagtlgaatgc 835
Db	820 AATTCACTGTGACAGTGTCCACATTCAGGGGtGTGTGGCTGTGTGGGtGTGGCTGTG 879
OY	836 aagcatcttgataacccgttacgtccaaactgttcggaactcgacgacgaatcttccaac 895
Db	880 AAGCCATCTCTGAGACAGGGCACCCTCCAAAGCTGTGGGGCCAGCAGACATCTCTAANA 939
OY	896 tttagcaagctatctggaagcacagaacagtlacogtltgaatttgacatagattgcaca 955
Db	940 tccagtagccCACTTTGGAGCCACACAGAAACCAATAGCATAGTTTACATCGACTCGTGCAGA 999



|||||  
Db 1253 ATGCCAGAAATGATACCTGCGGGATGTTTCATCCGAGATATACGCGCTTTGACA 1312  
Qy 1136 gggccaaacacccctggtggtcgtacgaagcaatctga 1173  
|||||  
Db 1313 GGGCCAAACACTGCGTGGGGCTGGCCAAAGCATCTGA 1350  
RESULT 6  
Al5836  
LOCUS Al5836 1291 bp DNA linear PAT 16-FEB-1994  
DEFINITION chymosin gene.  
ACCESSION Al5836  
VERSION Al5836.1 GI:488959  
KEYWORDS  
SOURCE .  
ORGANISM cow.  
Bos taurus  
Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovidae; Bovinae; Bos.  
REFERENCE  
1 (bases 1 to 1291)  
Simons,A.F.M. and De Vos,W.M.  
DNA fragments, containing a lactic acid bacterium-specific  
regulator region for the expression of genes coding for normally  
heterologous proteins  
Patent: EP 0307011-A 5 15-MAR-1989;  
JOURNAL NEDERLANDS INSTITUUT VOOR ZUIVELONDERZOEK  
FEATURES  
source 1..1291  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
BASE COUNT 323 a 382 c 328 g 257 t 1 others  
ORIGIN  
Query Match 76.4%; Score 896.2; DB 6; Length 1291;  
Best Local Similarity 87.6%; Pred. No. 9.5e-262;  
Matches 979; Conservative 0; Mismatches 139; Indels 0; Gaps 0;  
Qy 56 tcgtgctgttactacacgtcgtatagatcacccgcatcctctctacaagaagtaagtctc 115  
|||||  
Db 9 TCTTCGCTCTCTCCAGGGGCTGAGATCACCAGATCCCTCTGTACAAAGCAAGTCTC 68  
Qy 116 tcgtgaagcgtctgaagaaatgactctagaagactctcttcgaagaacaagatg 175  
|||||  
Db 69 TGAGGAAGCGCCTGAGAGAGATGGGCTTGTGAGAGACTTCTCGAAGAAAGCAGTATG 128  
Qy 176 gcatcagcagaagtaactcggcttgggtgaagtgtcagtgctgacattaccactacc 235  
|||||  
Db 129 GCATCAGCAGCAAGTACTCCGGCTTGGGAGGTGGCCAGCGTGCCTTGACCAACTTACC 188  
Qy 236 ttgataagcaatcttgggaagaatctaccccgaaaccgcctcaagaagtctacgcttc 295  
|||||  
Db 189 TGGATGATGACTACTTGTGGAGATCTACCTCGGAGACCCGCCCAAGAGTTCAACCGTGC 248  
Qy 296 tcttgataactggtctcctctactcttgggttccctctatactactcaagagcaatgctc 355  
|||||  
Db 249 TGTGTGACACAGGCGCTCTGACTTCTGGGTATCCTTATCTATGCAAGACATGCTCT 308  
Qy 356 gcaagaaccacaagaattcgatccgagaagaatgcttcaaccttcagaacttaagcaaac 415  
|||||  
Db 309 GCAAAAACCCAGCGCTTCGACCCGAGAAAGTCTCCACCTTCAGAACTGGGCAAGC 368  
Qy 416 ccttgtatatactagcttagcaggttagcatcaaggaattcttagctatgatacctga 475  
|||||  
Db 369 CCTGTCTATCTACTACGAGACAGCAGCATCAAGGCAATCTCTGGGCTATGACACCGTCA 428  
Qy 476 ctgtctccaacatctgtgacatctcaacagacagtaggaacttagcaccacaagaacagtg 535  
|||||  
Db 429 CTGTCTCAACATTTGTGGACATTCAGACAGTAGAGCTTGACACCCGAGGCCGGGG 488  
Qy 536 atgtcttcaactatgcaagaattcgatgctccttggtatggtgacatccatgcctgctg 595  
|||||

Db 489 AGCTTTACCTTATGCCGAATTGACGGGATCTCTGGGGATGGGCTACCCCTCGCTGGCT 548  
Qy 596 cagagtaactcgaatacctgtgttgacaacaatga tgaaccgacacctaagtaagtaagta 655  
|||||  
Db 549 CAGAGTACTCGATACCCGCTTTTGACAAATGATGAACAGGACACCTGTTGGCCCAAGACC 608  
Qy 656 tgtcttcgtttacatggaagaagaatgagcagaagagcagctcagaagctggagctatg 715  
|||||  
Db 609 TGTTCGTGGTTTACATGAGACAGGAATGGCCGAGAGACATCTACGCTTGCGGCCATCG 668  
Qy 716 atccatcctactacaagaatctccttcaactgggttccagcaactgtgcaagtaagta 775  
|||||  
Db 669 ACCONCTTACTACACAGGGTCTCTGACTGCTGGTCCCGGTGACAGTTCACAGTACTGCG 728  
Qy 776 aattcactgtgacagtgatccatcagcagtggtgtgtgtgtgtgtgtgtgtgtgtgtgt 835  
|||||  
Db 729 AGTTCATCTGTGACAGTGTACAGATCAGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 788  
Qy 836 aagctatcttgataaccggtacgttcaagctgtgtcgaacttagcagagacatctcagac 895  
|||||  
Db 789 AGGCAATCTTGACACGGGACCTCCAAAGCTGTGTGGGCCAGACGACATCTCTACACA 848  
Qy 896 ttccagaagctatgtgagcacaacagacagtaagtgatgtgatatagatgtgcagac 955  
|||||  
Db 849 TCCAGCAGGCCATTTGAGACCAACAGACAGACAGTACGCTGATGTGACATGACAGTGCACA 908  
Qy 956 accttaactatagctcactacagtgcttcttgatgaacaagcagaatgtaccactgaacc 1015  
|||||  
Db 909 ACCTGACTATACATGCCCACTGT 968  
Qy 1016 ccttcgctatcacagcagaagatcaagtgctctcaccagtgatgttcagaagtaagacc 1075  
|||||  
Db 969 CCTCCGCTATACCAAGCAGGACCAAGGCTCTCTTACCAAGTGGCTTCCAGAGTAAATATC 1028  
Qy 1076 attccagaagaatgatactgtgagagatgtgtcattcgtgagtaactacagcgtcttgaca 1135  
|||||  
Db 1029 ATTCGCAAAATGATGATCTGTGGGGATGTTTCATCCGAGATATACAGCGCTTTTGACA 1088  
Qy 1136 gggccaaacacccctggtggtcgtacgaagcaatctga 1173  
|||||  
Db 1089 GGGCCAAACACTGCTGTGGGCTGGCCAAAGCATCTGA 1126  
RESULT 7  
I04058  
LOCUS I04058 1175 bp linear PAT 02-DEC-1994  
DEFINITION Sequence 5 from Patent EP 0123928.  
ACCESSION I04058  
VERSION I04058.1 GI:591912  
KEYWORDS  
SOURCE .  
ORGANISM Unknown.  
Unclassified.  
REFERENCE 1 (bases 1 to 1175)  
AUTHORS Cashion,L.M., McCaman,M.T., Rice,C.W. and Sias,S.R.  
TITLE Recombinant DNA coding for a polypeptide displaying milk clotting  
activity  
JOURNAL Patent: EP 0123928-A2 5 07-NOV-1984;  
FEATURES  
source 1..1175  
/organism="Unknown"  
BASE COUNT 263 a 356 c 318 g 238 t  
ORIGIN  
Query Match 76.4%; Score 895.6; DB 6; Length 1175;  
Best Local Similarity 87.6%; Pred. No. 1.4e-261;  
Matches 979; Conservative 0; Mismatches 139; Indels 0; Gaps 0;  
Qy 56 tcgtgctgttactacacgtcgttgaatcacccgcatcctctctacaagaagtaagtctc 115  
|||||  
Db 49 TCTTCGCTCTCTCCAGGGGCTGAGATCACAGATCCCTCTGTATCAAAAGCAAGTCTC 108

OY	116	ttgtatgaacgcgttgaggaaacatgagactcttagaagactctcttcgaaacaaacatg	175
Db	109	ttgaggaagcgccttgaaagagcatgagggctcttgaggagactcttcgcaaaaacacatg	168
OY	176	gcatcaagcagaagtaactcgcgcgttcggttgaaagtgtgtagcgtgtccacttaacatacc	235
Db	169	gcattcagcagcaagtaactccggcttccggggagagtgggcaccagctggtcccttaccacattacc	228
OY	236	ttgatatgtcaataacttttggagaatctactcgtgaaccgcgcctcaagattcaacgttc	295
Db	229	tggaaatgtaagtaacttgggaagatctatccctgggaccctccggacagattcaccgtgc	288
OY	296	tcttttgaactagtggttcctctgaactcttgggttcctctatctactcgaagaagaatgact	355
Db	289	tgttttgacatcgtgcctcctcttactcttgggttaccctcttatcttctgcamaagcaatgcct	348
OY	356	gcaagaaccacacaaagattcagatccgaagaagtctgcactcttcagaactttagcaaac	415
Db	349	gcaaaaaacacacagcggcttccaccggagaaagtgtccaccttccagAACCTGGCAAG	408
OY	416	ccttgtctatcacactacagtgatacagttatgtcaagaagaatcttagctatgatacgtta	475
Db	409	ccctctctatccactacacgggacagcagcagcagcagggcattcmaagcctTAGAACCCGTA	468
OY	476	ctgtcccaacaatltgtgacatcaacagaacagtaagacttagcacccaagaacagagtg	535
Db	469	ctgtctccAACATTGTGGACATCCAGAGAGACAGTACGCTTGACACCCAGAGACCCGGGG	528
OY	536	atgtcttcaactatgcagaatctgaatgagcalccttgtagtataccaaccatgcgtcgt	595
Db	539	acgtctttcaccattatgcccaatttcacacgggattccggggagtgggccttacccttcgtccgt	588
OY	596	cagagtaactgataacgttgtgtgttacaacaatgataacgcaacacctaagtaagtcacagat	655
Db	589	cagaattactcattatcccgctgtttttacAACATGATGAACAGGCAccttggtggccCAAGACC	648
OY	656	tgttcctggttacaatgtagacaggaatgtgcacagaagagacatgctccacagctltgagctatgt	715
Db	649	tgtttccggtttTAATATGAGACAGGAATGACCAGAGACATGCTCACGtgggggCCATGC	708
OY	716	atccatcctactaaacaggaatctctcttaacgtgttccagactatgtgcagcagtaactgtgc	775
Db	709	accctgtccttacttACACAGGGGCTCGTGAAGTGGGCCCCGTGAACAGTGCACAGTACTGGCC	768
OY	776	aattcaactgttgagacagtgatccacatcacagcgtgtgtgtgtgtcatgttgaagtgtgatgc	835
Db	769	agttactcgtgagacagtgctacatfcaacgtacggcgtgtgtgtggccttgaagggtgcgtgtc	828
OY	836	aagctatcttggataccggtatcgttccaagcttgttcggaactagcagcagatctcaaca	895
Db	839	agggcattccttgagAACACGGGCAccttccAAAGTGTGCGGGCCACACAGCATCCTCAACA	888
OY	886	ttcagaacagcatltggagccacacagaaccagtaagtggtgtgtgtgataagattgtagata	955
Db	889	tgcacacagccatttggaagccacacAGAACAGTACAGTATGATTGACATCAGATGCGGACA	948
OY	956	accattagctacatgcctactaagtggtctcttggatgtcaacagggcaagaatgtaccactgtacc	1015
Db	949	acctttagcttactATCCCACTGTGTCTTTGAGATCAATGTGCAAAATGTACCACACTGACC	1000
OY	1016	cctgcgcctataccagccagatataaaggttctgcacaccagttgattccagaagtgtgaac	1077
Db	1009	cctgcgccttATACCAAGCCAGAACACAGGGGCTTGTGTACCAATGTGCTTCCAGAGTGAATAAC	1066
OY	1076	attccacgaatgtgactcttggagatgtgtgtcatctcgtgtgactacagcgtctgtgaca	1135
Db	1069	attccacgaatgtgactcctgcgggatgttttcaatccgagagatgtattACACGtctttTGACA	1128
OY	1136	ggggcacaacactcgtgtgggtatgataagaacatctga	1173
Db	1129	ggggcacaacactcgtgtggggctggccAAACCACTGTGA	1166

RESULT	8			
AR073077				
LOCUS	AR073077	2733 bp	DNA	linear
DEFINITION	Sequence 6 from patent US 5948682.			PAT 28-AUG-2000
ACCESSION	AR073077			
VERSION	AR073077.1	GI:9999840		
KEYWORDS	.			
SOURCE	Unknown.			
ORGANISM	Unclassified.			
REFERENCE	1 (bases 1 to 2733)			
AUTHORS	Moloney,M.M.			
TITLE	Preparation of heterologous proteins on oil bodies			
JOURNAL	Patent: US 5948682-A 6 07-SEP-1999;			
FEATURES	Location/Qualifiers			
SOURCE	1..2733	/organism="unknown"		
BASE COUNT	699 a	681 c	684 g	669 t
ORIGIN				

Query Match	76.3% ;	Score 894.8 ;	DB 6 ;	Length 2733 ;
Best Local Similarity	88.4% ;	Pred. No. 2.7e-261 ;		
Matches 971 ;	Conservative 0 ;	Mismatches 127 ;	Indels 0 ;	Gaps 0 ;
QY 76	gctgagatcacccgcgatctctctctactacaagaagtaagtcctccgtaagcgctgaagaa	135		
Db 1630	gctgagatcacccgcgatctctctctactacaagaagtaagtcctccgtaagaaagcgctgaagaa	1689		
QY 136	catggactcttaagaagactctctgcagaaacacagatattggatcagcaagaagttacc	195		
Db 1690	catggactcttctggagagacttctctcagaaacacagatattggatcagcaagaagttacc	1749		
QY 196	ggcttcggtgaagttgctcagcgctgcacactaacactaccttgaagtaagtaattctggg	255		
Db 1750	ggcttcggtgaagttgctcagcgctgcacactaacactaccttgaagtaagtaattctggg	1809		
QY 256	aagatctacccctcgcgaaccccgctccaaagttcaccgcttctccttgaacgtgtccctc	315		
Db 1810	aagatctacccctcgcgaaccccgctccaaagttcaccgcttctccttgaacgtgtccctc	1869		
QY 316	gaactctgggtgtccctctatctactgcgaagagcaatgcctgcgaagaaaccaaaagatc	375		
Db 1870	gaactctgggtgtccctctatctactgcgaagagcaatgcctgcgaagaaaccaaaagatc	1929		
QY 376	gattccggaagaagtgtgtccactcttcacaaacttaagcaaaacctgtgtataaactaagtc	435		
Db 1930	gattccggaagaagtgtgtccactcttcacaaacttaagcaaaacctgtgtataaactaagtc	1989		
QY 436	acaaagtacatgcagaaggaatctttagctatgatataccgtacactgtctccaaatgtgac	495		
Db 1990	acaaagtacatgcagaaggaatctttagctatgatataccgtacactgtctccaaatgtgac	2049		
QY 496	attcaacacagactagaacttaagcaccacaaagaaccaagtgatgtcttcaactatgcaga	555		
Db 2050	attcaacacagactagaacttaagcaccacaaagaaccaagtgatgtcttcaactatgcaga	2109		
QY 556	ttcgatgtgcattccttggatattgacataccactgcgtctgcagaagtacttgataactgtg	615		
Db 2110	ttcgatgtgcattccttggatattgacataccactgcgtctgcagaagtacttgataactgtg	2169		
QY 616	ttctacacacatgatgaacccgcacacttagttagctaaagaactgttctcgtgttacaatgac	675		
Db 2170	tttgcacacacatgatgaacccgcacacttagttagctaaagaactgttctcgtgttacaatgac	2229		
QY 676	aggaatgycacagaagacatgctcagcgtcttgaaagtctatgtatccatccactactacaaga	735		
Db 2230	aggaatgycacagaagacatgctcagcgtcttgaaagtctatgtatccatccactactacaaga	2289		
QY 736	tctcttcaactgggtgtcctcagtcactgtgtgcagcagtaactgtgcgaattcactgtgtgacagtgc	795		
Db 2290	ttcccttgacatgggtgtcctcagtcactgtgtgcagcagtaactgtgtgacagtgtgc 2349			



QY	DB	1076	attccagaatggaatctgggagatgcttcattcttgtagtactacaagctcttgaca	1135	
QY	1016	ccctcgccatataccagcagcagatcaagggttctgtcacccagtgatctccagatgagaacc	1075		
DB	1009	ccctccgccctatpaccacacccagacacagggcttctgtaccagtgcttccagagtaaaac	1068		
QY	1076	attccagaatggaatctgggagatgcttcattcttgtagtactacaagctcttgaca	1135		
DB	1069	attccacagaaatgagatccctggggagatgcttcttcattccagagatgtaacagcgtcttgaca	1128		
QY	1136	ggggcaacaacctgctgggagatgcttcattcttgtagtactacaagctcttgaca	1173		
DB	1129	ggggcaacaacctgctgggagatgcttcttcattccagagatgtaacagcgtcttgaca	1166		
RESULT	10				
E00075					
LOCUS	E00075	1290 bp	RNA	linear	
DEFINITION	CDNA encoding chymosin ("rennin").				
ACCESSION	E00075				
VERSION	E00075.1 GI:2168379				
KEYWORDS	JP 1983009687-A/1.				
SOURCE	cow.				
ORGANISM	Bos taurus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea; Bovidae; Bovinae; Bos.				
AUTHORS	1 (bases 1 to 1290) Noonan,H.K., Malkeru,T.D., Chlmoshil,J.R.H.H., Pilta,A.R. and Jiyon,S.E.				
TITLE	PRODUCTION OF POLYPEPTIDE				
JOURNAL	Patent: JP 1983009687-A.1 20-JAN-1983;				
COMMENT	CELLTECH LTD OS PN JP 1983009687-A/1 PD 20-JAN-1983 PF 17-JUN-1982 JP 1982104672 PR 17-JUN-1981 GB 81 8118688, 11-NOV-1981 GB 81 8133998, PR 01-DEC-1981 GB 81 8136185, 10-FEB-1982 GB 82 8203907 PI NOOMAN HERRI KEARI, MAIKERU TERENSU DOORU, PI CHIMOSHIL JIYON ROI HARIISU, PI CHIMOSHIL JIYON ROI HARIISU, PI PIITA ANSONIT ROU, JIYON SUPENSAA EMUTPEJL PC C12N9/52.C07H21/04.C12N1/00.C12N9/60.C12N15/00, PC C12N1/02/C12R1/19, PC C12R1/865; CC strandedness: Double; CC topology: linear; CC hypothetical: No; CC *source: tissue.type=stomach; CC *source: clone-pct 70; FH key FH Location/Qualifiers FT 5'UTR 1..25 FT sig_peptide 26..73 FT mat_peptide 74..1078 FT /product='pro-chymosin' FT 3'UTR 1082..1200 FT CDS 26..1081 FT /product='pre-pro-chymosin'. FEATURES source Location/Qualifiers 1..1290 /organism="Bos taurus" /db_xref="taxon:9913" BASE COUNT 309 a 384 c 334 g 263 t ORIGIN				
Query Match	76.2%:	Score 894;	DB 6;	Length 1290;	
Best Local Similarity	87.5%:	Pred. No. 4,4e-261;			
Matches 978:	Conservative 0;	Mismatches 140;	Indels 0;	Gaps 0;	
56	tcgttcgcttactcaacgctctgagatcacccgcattctctctacaagaagttaagcttc	115			

[illegible]

RESULT 11  
LOCUS A15633 1291 bp mRNA linear PAT 18-FEB-1994  
DEFINITION preprochymosin.  
ACCESSION A15633  
VERSION A15633.1 GI:491951  
KEYWORDS  
SOURCE synthetic construct.  
ORGANISM artificial sequence.  
REFERENCE 1 (bases 1 to 1291)  
AUTHORS Carey,N.H., Doel,M.T., Harris,T.J.R., Lowe,P.A. and Emlage,J.S.  
TITLE A process for the production of a polypeptide  
JOURNAL Patient: EP 0068691-A 29 05-JAN-1983;  
CELLTECH LIMITED  
FEATURES  
source Location/Qualifiers  
1..1291  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
26..1171  
/gene="preprochymosin"  
26..1171  
/gene="preprochymosin"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAA01244.1"  
/db\_xref="GI:491952"  
/translation="MRCLVYLLAVFALSGAETTRIPLYKSKLRKLRKHGLEDFL  
OKQOYISKYSSEGFVAVPLTNVLDOSITFGKIVGATPEQETFLVDGSSPWFPS  
YCKSNACKNHQRFDPKRSSTFQNLGKPLSIHGTGSMOILIDYDVYNSIVDIOQT  
VGLSTGEQDVYFAEFDFILGMAYPSLSEVSIPEVDNMNHLVAODLFSVMORI  
GQSMETLGAINPSYTGSLHWPVTVQOYFQVDSVTISGVVAECGCGQAILDTG  
TSLTVBPSSDILNIOALIGATONOGEPFDICDNLSPVTVFEINKMPLPSAVT  
SDQGFCTSGFSGSEHNSOKNITLDFVIREYVSPFDANNLVGLAKAI"  
misc\_difference 1291  
BASE COUNT 309 a 384 c 334 g 263 t 1 others  
ORIGIN  
Query Match 76.28; Score 894; DB 6; Length 1291;  
Best Local Similarity 87.5%; Pred. No. 4.4e-261;  
Matches 978; Conservative 0; Mismatches 140; Indels 0; Gaps 0;  
QY 56 tcattgtctgttaccagcgtctagatcacccgcatccctctacaaagtaagcttc 115  
Db 54 TCTTGGCTCTCCCAAGGCGCTGAGATCACAGAGATCCCTGTACAAAGCAAGTCTC 113  
QY 116 tccgtaaggcgtgaaggaacatggaactctagaagaactcttgcaagaacaagatag 175  
Db 114 TGAGGAAGGCGCTGAGAGAGATGGGCTTGTGAGAGACTTCTGCAAGAACAGCAGTATG 173  
QY 176 gcatcagaagaagtactcggctcggttggaagtgtcgaagtgccactaccactcc 235  
Db 174 GCATCAGCAGCAAGTACTCCGGCTTGGGAGGTGGCCAGCGTCCCTGACCAACTATCC 233  
QY 236 ttgataagtaactcttggaagaactaccccggaaccccgccaaagaatcgaagcttc 295  
Db 234 TAGATAGTCACTACTTTGGGAAGATCTACTCGGAGACCCGCCCCAGAGATTTACCGGTGC 293  
QY 296 tctttgatactggtctcctctgactcttggtctccctctatactactgcaagaatgctc 355  
Db 294 TGTTTGACACTGGGCTCCTGACTTCTGGTATACCTCTATCTACTGCAAGCAATGCTCT 353  
QY 356 gcaagaaccaccaagatcgcagagaagtgctcactctccagaacttaggcaaac 415  
Db 354 GCAAAAACACACAGCGCTTGCACCCGAGAAAGTCTCCACCTCCAGAACTTGGCGAAGC 413  
QY 416 cctgtctatactactgaagtagcatgacgaagaagaactcttagctatatactacgtca 475  
Db 414 CCCTGTCAATCACTACGGGACAGGACAGATGCAAGGCAATCTTAGGATATGACACCGTTA 473

QY 476 ctgtctccaaacatttgygacatcacaagacagtlaggaacttagcaaccaagaacagtg 535  
Db 474 CTGTCTCCAAACATTTGTGACATCCAGACAGACAGTAGGCTGACACCCAGAGCCGGG 533  
QY 536 atgtcttcaacctatgcagaatctgagatgcactctcttgtaggaataccatctgcgtc 595  
Db 534 ACGTCTTCACTATGCGGAATTCGACGGAATCCTGGGGAGGCTTACCTCCCTGCTGCT 593  
QY 596 cagaagtaactgataccctgtgttgtagacacatgataacacacactagtagtaaacact 655  
Db 594 CAGAGTACTGATACCCGTGTTTGACAAACATGATGAAACAGGCAATTTGGTGGCCAAAC 653  
QY 656 tgtctcggttatactatgacaagaatgagccagaagaagatgctcgaagcttgagactatg 715  
Db 654 TGTCTCGGTTTACATGACAGATGATGGTCAGAGAGACATGCTCACCTCGGGGCGCATCA 713  
QY 716 atccatcctatacaagaagatctcttactcgtggtctcgaagcactgctgcaagaactg 775  
Db 714 ACCCGTCTACTACACAGGGTCCCTGCACTGGGTGCCGTGACATGTCAGAGTACTGCC 773  
QY 776 aattcactgtgagacgtgtcaaccatcagcgtgtgtgtgtgtgtcatgtgaagtgatgtc 835  
Db 774 AGTTCACTGTGGACAGTGTACACATCAGCGGTGTGTGATGCGCTGTGAGAGGTGGTGC 833  
QY 836 aagctatcttggataccggtacgtlccaagctgtgcgaacctagcagcgaacttccaaca 895  
Db 834 AGGCATCTTGGACACGGGGCACTCCAAAGCTGTCGGGGCCAGCAGGACATCTCCACAA 893  
QY 896 ttgacgaagctatggagacacacagaagaacagatgcgttgagttgataagatgagaca 955  
Db 894 TCCAGCAGGCCATTGGAGGCCACACAGAACAGTACGGGTGATTTGACATGCAGTCCGACA 953  
QY 956 accttaagctatgcctcactacagttgtcttgagatacaacggaagatgtaaccactgacc 1015  
Db 954 ACCGAGGTATATCCCACTGTGCTTGTGATGATCAATGGCAAAATGATACCCACTGACC 1013  
QY 1016 cctcgcctatcacagccagaatcaagggtctcgcaccagtgatgtccaagatgagaacc 1075  
Db 1014 CCTCCGCTATACCAAGCCAAACCAAGGCTCTGTACCAAGTGGCTTCCAGAGTAAATATC 1073  
QY 1076 attccagaagaatgagatcttggagaaatgtgtcattctcgttgagtaactaagcgtttgaca 1135  
Db 1074 ATTCCCAAAATGGAATCTGGGGAGTGTTCATCCCGAGATATATTACAGCGCTTTGACA 1133  
QY 1136 ggccaacaacctcgttgggttagctaaagaactctga 1173  
Db 1134 GGGCAACAACCTCGTGGGGCTGGCCAAAGCCATCTGA 1171  
RESULT 12  
LOCUS E00144 1289 bp RNA linear PAT 29-SEP-1997  
DEFINITION cDNA encoding bovine calf chymosin.  
ACCESSION E00144  
VERSION E00144.1 GI:2168443  
KEYWORDS JP 1984021392-A/1.  
SOURCE unidentified.  
ORGANISM unclassified.  
REFERENCE 1 (bases 1 to 1289)  
AUTHORS Chiyarrusu,E.B.  
TITLE CATTLE KIMOCIN  
JOURNAL Patent: JP 1984021392-A 1 03-FEB-1984;  
GENEX CORP  
COMMENT OS bovine calf  
PN JP 1984021392-A/1  
PD 03-FEB-1984  
PF 30-JUN-1983 JP 1983119481  
PR 01-JUL-1982 US 82 394433. 13-APR-1983 US 83 484539 PI  
CHYARRUSU EI BASURETSUTO  
PC C12N15/00, C07H21/04, C12N1/20, C12P19/34, C12P21/02, (C12N15/00,  
C12R1:19);







Db 532 CAGAGTACTCGATACCCGTTGTCAGAACATGATGAACAGGACCTGGTGGCCCAAGACC 651  
Oy 656 tcttcggtttacatgacaggaatggccagagagacatgctcaagcttgagctatg 715  
Db 652 TGTTCGCTTACATGACAGGAATGGCCAGAGACCATGCTCAAGCTGGGGGCATG 711  
Oy 716 atccatccactacacagagatctctcactggttccagctcactgctgacagactgac 775  
Db 712 ACCGCTCTACTACACAGAGGTCCTGCAATGGGTGCCGTGACAGTGCAGCAGTACTGGC 771  
Oy 776 aattcaactgtggacagtgctcaccatcaagcgtgtgtgtgtgtgtgtgtgtgtgtgt 835  
Db 772 AGTTCAGTGTGACAGTGTACCATCAGCGGCTGTGTGTGTGTGTGTGTGTGTGTGTGT 831  
Oy 836 aagcatcttgatccggttgaagctgctgaagctgtgtgtgtgtgtgtgtgtgtgtgtgt 895  
Db 832 AGGCAATCTGTGACAGGAGGACCTCAAGCTGTGTGTGTGTGTGTGTGTGTGTGTGT 891  
Oy 896 ttcagcaagctatgtgagccacacagaaacagtaagctgagcttgatgacatgtgcgaca 955  
Db 892 TCCACAGAGCCATTGGAGCCACACAGAACCACTAGCTGATGATTGACATCGACTGCGACA 951  
Oy 956 accatgactacatgcctcactgagtgcttcttgatgacacgagcaagatgtaccactgacc 1015  
Db 952 ACTGAGCTACATGCCCTGT 1011  
Oy 1016 cctcggcctatccagcagagatcaaggttctgacacagctgagcttccaagctgagacc 1075  
Db 1012 CCTCGGCTATACACAGGACGAGGAGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1071  
Oy 1076 attcccaaatgagctcttggagatgtgtcattcgttgagctgagctgagcttggaca 1135  
Db 1072 ATTCACCAAAATGATCTCTGGGGGATGTTTCATCCGAGAGATATACACGCTTTGACA 1131  
Oy 1136 gggccaacaacctcgttggctagctaaagaacttga 1173  
Db 1132 GGGCCAAACACCTCGTGGGGCTGGCCAAAGCCATCTGA 1169

RESULT 15  
LOCUS E00295 1460 bp RNA linear PAT 29-SEP-1997  
DEFINITION cDNA encoding pre-prorenin A.  
ACCESSION E00295  
VERSION E00295.1 GI:2168583  
KEYWORDS JP 1985058077-A/3.  
SOURCE cow.  
ORGANISM Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovidae; Bovinae; Bos.  
1 (bases 1 to 1460)  
Deblutsudo,B., Konarudo,U.D., Jierarudo,R.F., Arison,T.R.,  
Rodaato,J.N., Jien,I.M., Donarudo,T.M. and Kurisutofuaa,G.G.  
USE OF GAL YEAST PROMOTOR  
Patent: JP 1985058077-A 3 04-APR-1985;  
KORABORATEIBU RES INC  
COMMENT  
TITLE JOURNAL  
AUTHORS  
REFERENCE  
OS Bovine  
PN JP 1985058077-A/3  
PD 04-APR-1985  
PF 28-FEB-1984 JP 1984035472  
PR 28-FEB-1983 US 83 470911  
PI DEBUTSUDO BOTSUTOSUREIN, ROMARUDO UEIN DEIBISU, PI  
JIERARUDO RARUFU FUINKU,  
PI ARISON TAUNTON RIBUBIT, ROBARATO JIENTORIT NOURUTON, JIEN I  
MAO, PI DONARUDO TEIRAA MOA, KURISUTOFUA GOTSUDOFURIT GOFU PC  
C12N15/00,C07H21/04,C07K13/00,C12N1/16,C12P21/02,C12N1/16, PC  
C12R1.865),  
PC (C12P21/02,C12R1.865);  
CC strandedness: Double;  
CC topology: Linear;  
CC hypothetical: No;  
CC anti-sense: No;

CC \*source: tissue type=Stomach mucosa;  
CC \*source: clone=293-207 & 293-118/37;  
FH Key Location/Qualifiers  
FH CDS 205..1350  
FT /product=/'pre-prorenin A'  
FT mat.peptide 205..1347  
FT /product=/'pre-prorenin A'  
FT mat.peptide 253..1347  
FT /product=/'Prorenin'.  
FEATURES  
source Location/Qualifiers  
1..1460  
/organism='Bos taurus'  
/db\_xref='taxon:9913'  
BASE COUNT 327 a 437 c 400 g 296 t  
ORIGIN  
Query Match 76.1%; Score 892.4; DB 6; Length 1460;  
Best Local Similarity 87.4%; Pred. No.1,4e-260;  
Matches 977; Conservative 0; Mismatches 141; Indels 0; Gaps 0;  
Oy 56 tctgttgcgttactacagctgctgagatcacccgcatctctctcaaaagtgtaagctc 115  
Db 233 TCTTCGCTCTTCCACAGGCGCTGAGATCACAGATCCCTGTGATCAAGCAAGTCTC 292  
Oy 116 tccgtgaagcgttgaaagaaatgtgactctgaagactctcttgcagaacaagatg 175  
Db 293 TGAGGAAGCGGTGAAGGAGCATGGGCTTCTGAGACACTCTGCAGAAACAGCACTATG 352  
Oy 176 gcatcagcagcaagctacccgcttcggtgaaagtgtgacgtgtgcaacttaccactacc 235  
Db 353 GCATCAGCAGCAAGTACTCCGGCTTGCGGACAGTGGCAGCGCTGCCCTTGACCACTACC 412  
Oy 236 ttgatagtacatcttggaaagatctacctggaaccccgccctcaagaagtctacagctc 295  
Db 413 TCGATATGTCAGTACTTGGGAAGATCTCGGGAGCCCCGCCGCGGAGTTACCGTGC 472  
Oy 296 tctttgatactggttcccttgacttgggttccctctctctactgcaagagaatgct 355  
Db 473 TGTGTGACACTGGCTCCTGTGACTCTTGAGTCCCTGTATCTGCAAGACCAATGCT 532  
Oy 356 gaaagaaccaccaagaatctgcagagaagctgtccacttccagaaacttaagcaaac 415  
Db 533 GCAAAAACCCACAGGCTTCGACCCGAGAAATGTCACCTTCCAGAACTGGGCAAC 592  
Oy 416 cctgtctctatacactacaggttagcagtagcagaaatcttagctatgatacagctca 475  
Db 593 CCTGTCTATTCACATACAGGACAGGACAGGATGCTGGGGCTATGATACCGTCA 652  
Oy 476 cgtgtctcaaatgttgacatccaacagagatgagacttagcaccaccaagaacagtg 535  
Db 653 CTGTCTCCAACTTGTGACATCCAGACAGCATGAGGCTGTGACACCCAGCGGGG 712  
Oy 536 atgtctcaacctatgcaagatctgcagatccctgtgtgtgtgtgtgtgtgtgtgtgtgt 595  
Db 713 AGCTCTTACCTATGCGCAATTCGAGAGGATCTGTGGGATGAGCCCTTGCTGCT 772  
Oy 596 cagagactacatgactgctgttggacaacatgataacacagcaactagtagctcaagact 655  
Db 773 CAGAGTACTCGATACCCGTTGTGACAAACATGATGAACAGCACTGGTGGCCAAAGAC 832  
Oy 656 tcttcggtttacatgacaggaatggccagagagacatgctcaagcttgagctatg 715  
Db 833 TGTTCGCTTACATGACAGGAATGGCCAGAGCATGCTGCAAGCTGGGGGCATCG 892  
Oy 716 atccatccactacacagagatctctcactggttccagctcactgtgcagcagctagct 775  
Db 893 ACCGCTCTACTACACAGAGGTCCTGCAATGGGTGCCGTGACAGTGCAGCAGTACTGGC 952  
Oy 776 aattcaactgtggacagtgctcaccatcaagcgtgtgtgtgtgtgtgtgtgtgtgtgt 835  
Db 953 AGTTCAGTGTGACAGTGTACACTTCACGCGGCTGTGTGTGTGTGTGTGTGTGTGTGTGT 1012





xx 01-MAR-2001.  
xx  
xx 23-AUG-2000; 2000MO-CA00975.  
xx  
xx 23-AUG-1999; 99US-0378696.  
xx  
xx (SEMB-) SEMBIOSYS GENETICS INC.  
xx  
xx Van Rooijen G, Keon RG, Boothe J, Shen Y;  
xx  
xx WP1: 2001-226621/23.  
xx DR P-PSDB; AAU00536.  
xx  
xx Producing chymosin in seeds of plants such as rice, flax, rape seed, by  
xx transforming plant cell with a nucleic acid encoding chymosin operably  
xx linked to transcription regulator and terminator sequences -  
xx  
xx Claim 9; Fig 1; 56pp; English.

The sequence represents a DNA which encodes a bovine chymosin polypeptide. Chymosin can be produced in a plant seed through introduction of a chimeric nucleic acid molecule, comprising a nucleic acid sequence encoding a chymosin polypeptide operatively linked to CC transcription regulator and terminator sequences, into a plant cell. The CC soybean are useful for producing plant seeds, in particular seeds of CC barley, oats, sorghum, Arabidopsis thaliana, potato, flax/linsed, CC safflower, oil palm, groundnut, Brazil nut, coconut, castor, coriander, CC squash, jojoba and rice.

Sequence 1173 BP; 299 A; 308 C; 262 G; 304 T; 0 other;

	Query Match	100.0%;	Score 1173;	DB 22;	Length 1173;
	Best Local Similarity	100.0%;	Pred. No. 0;		
	Matches 1173;	Conservative	0;	Mismatches	0;
				Indels	0;
				Gaps	0;
QY	1 atgaacttccttaagcttcctcccttccttaagcttccttcgttcgttcgcaatactcgtt	60			
Db	1 atgaacttccttaagcttcctcccttccttccttaagcttccttcgttcgttcgcaatactcgtt	60			
QY	61 gctgttactcaagctcgtctgtagatcacccgcaattcctctctatacaagaatgctctcgt	120			
Db	61 gctgttactcaagctcgtctgtagatcacccgcaattcctctctatacaagaatgctctcgt	120			
QY	121 aagcgcctggaaggaacatggaacttctaagaagctcttcgcaagaacaacatgagc	180			
Db	121 aagcgcctggaaggaacatggaacttctaagaagctcttcgcaagaacaacatgagc	180			
QY	181 agcagcaagtaactccggttcgtagaagtgtgtagcgtgcacattacaactacttgat	240			
Db	181 agcagcaagtaactccggttcgtagaagtgtgtagcgtgcacattacaactacttgat	240			
QY	241 agtcaatactttgggaagaatctactccctcggaaccccgctcaagaagtccaacgttctctt	300			
Db	241 agtcaatactttgggaagaatctactccctcggaaccccgctcaagaagtccaacgttctctt	300			
QY	301 gatactggttccctcgtactcttggttccctctatctatctgaagaacatgctctgaag	360			
Db	301 gatactggttccctcgtactcttggttccctctatctatctgaagaacatgctctgaag	360			
QY	361 aacacccaagaattcgatccggaagaagtcgtccactctccgaacttgggcaaaccttg	420			
Db	361 aacacccaagaattcgatccggaagaagtcgtccactctccgaacttgggcaaaccttg	420			
QY	421 tctatcacctacgtagacagtagacaatgaagaacattagagtagatcacgtctacgtc	480			
Db	421 tctatcacctacgtagacagtagacatgaagaacattagagtagatcacgtctacgtc	480			
QY	481 tccaacatctgtggaatccaacagaacagtagaacttgaoccccaagaacagtgatc	540			
Db	481 tccaacatctgtggaatccaacagaacagtagaacttgaoccccaagaacagtgatc	540			

OY	541	ttccctcatgcaagattcgatggtcagcttccttgatggtacccattcgctgcgtcaag	600
Db	541	ttccactatgcaaatctcgatggcatctccttggtatggtacaccacgcttcgcgtcaag	600
OY	601	tactcgaatacctggttttgacaacatgatgatgaacccagaccactagtagtctaaagactgttc	660
Db	601	tattcgcataactggttttgacaacatgatgatgaacccagaccactagtagtctaaagactgttc	660
OY	661	tcggtttacatgatgacaaggaatggccaaagagaacatgctctcaagcttgagagctattgatcca	720
Db	661	tcggtttacatgatgacaaggaatggccaaagagaacatgctctcaagcttgagagctattgatcca	720
OY	721	tcctactacaagaagatctcttcactcigtgttcctcagtgtaactggtgacgaagracgtggaattcc	780
Db	721	tcctactacaagaagatctcttcactcigtgttcctcagtgtaactggtgacgaagracgtggaattcc	780
OY	781	actgtgacaaatgtgcacacatccacagcgtgtgtgtgttcgtacatgtaaggttgatccaagct	840
Db	781	actgtgacaaatgtgcacacatccacagcgtgtgtgtgttcgtacatgtaaggttgatccaagct	840
OY	841	atctctggaataccggttacggtccaaagctgtgtcgcacctagacagcgacattctcaacattcaag	900
Db	841	atctctggaataccggttacggtccaaagctgtgtcgcacctagacagcgacattctcaacattcaag	900
OY	901	caagctatttgagagccacacagaacccagtaagctgagattggaatagattgacgaacaactt	960
Db	901	caagctatttgagagccacacagaacccagtaagctgagattggaatagattgacgaacaactt	960
OY	961	agctacatgccttaacagttcgtctcttgagatccaacggaagatgtaccacactgtgaccccttc	1020
Db	961	agctacatgccttaacagttcgtctcttgagatccaacggaagatgtaccacactgtgaccccttc	1020
OY	1021	gctctataccagccagatccaagggttctcgcacacagtgtgattccagaagtgagaacacttcc	1080
Db	1021	gctctataccagccagatccaagggttctcgcacacagtgtgattccagaagtgagaacacttcc	1080
OY	1081	cagaagaatggaatcttggaagaatggtgttcattctcgtgaatctactaaagcgtcttttgaacaggcc	1140
Db	1081	cagaagaatggaatcttggaagaatggtgttcattctcgtgaatctactaaagcgtcttttgaacaggcc	1140
OY	1141	aacaacctcgtttggtgcgtacgtaaaagcaatctga	1173
Db	1141	aacaacctcgtttggtgcgtacgtaaaagcaatctga	1173
RESULT 2			
AAS00570			
ID	AAS00570 standard; DNA; 3957 BP.		
XX	AAS00570;		
AC	1..11553		
XX	14-MAY-2001 (first entry)		
DT	Bovine phaseolin promoter pre-pro-chymosin-phaseolin terminator.		
DE	Chymosin: transcription regulator; terminator sequence; soybean; corn; pre-pro-chymosin; rape seed; sunflower; cotton; tobacco; alfalfa; wheat; barley; oats; sorghum; Arabidopsis thaliana; potato; flax; linseed; rice; safflower; oil palm; ground nut; Brazil nut; coconut; castor; coriander; squash; jojoba; ds; phaseolin; promoter; terminator; mutant; French bean		
XX	Chimeric - Bos sp.		
OS	Chimeric - Phaseolus vulgaris.		
XX	Key		
FX	Location/Qualifiers		
FT	1..11553		
FT	/*tag= a		
FT	/note= "Phaseolin promoter"		
FT	1154..2726		
FT	/*tag= b		
FT	/product= "Bovine pre-pro-chymosin"		
FT	2727..3957		
FT	terminator		







PS Claim 6; Fig 2; 39pp; English.

CC Also claimed is E.coli JM83/PLC7 (ATCC 39325) which is transformed  
CC with PLC7 contg. the prorenin derived sequence fused in phase with  
CC B-galactosidase. The PLC7 prorenin expression plasmid includes  
CC sequences which code for both the pseudorenin and mature rennin  
CC cleavage sites between Aas 28-29 and Aas 42-43, respectively.

XX  
Sequence 1175 BP; 263 A; 355 C; 319 G; 238 U; 0 other;

Query Match 76.5%; Score 897.2; DB 5; Length 1175;  
Best Local Similarity 68.2%; Pred. No. 7e-273;  
Matches 763; Conservative 217; Mismatches 138; Indels 0; Gaps 0;

```

OY 56 tcgttcgcttaccacgtctgtgagatccgacccgctcctcctacaaagttaagtctc 115
DB 49 uuuuugcucucucccaaggcgccgaggaucacccaagauccuuguaacaagcaagucuc 108
OY 116 tcgtaagcgctgaaggaacatgtgactctagaagactcttgtagaacaacagatag 175
DB 109 ugaggaagcgcgugagaagcgaugucuuugaggaucuuucugagaacaacagcaguanug 168
OY 176 gcatcagcagaagtaactcgcggtctgggtgaagtgtcagcgttcacttaccactacc 235
DB 169 gcaucagcagaaggaucuccgcuucggggagugugccagcgugcccuugaccauaccc 228
OY 236 ttgtagcacaatctcttggaagatctaccctcggaaccccgctcacaagttcacgcttc 295
DB 229 uggaugacgaucuuuugaggaucuaucucgggaccccgccccaaggaugacacgucg 288
OY 296 tcttgatactggttcctcctactctgtggttcctcctatcactatcgaagcaatgct 355
DB 289 uguuugacacugcucuccucugacuuucuguaaccuacuaucugcaaggaacugccu 348
OY 356 gcaagaaccaccaagatcgatccgagaagtcgtccaccttcacagaacttaggcaaac 415
DB 349 gcaaaaaaccacagcgcuucgacccgagagaagugucacacuuacagaaccugcgcaagc 408
OY 416 ccttgctatcactacgttagcagtagcatgcaaggaatcttagcgtatgatactgca 475
DB 409 cccuuguaucacacacagcggaaggaugcaugcaggaucuccuacugacacccgca 468
OY 476 ctgtccacaactgtgtgacatctcacagacagtagaagacttagccccaagaacagtg 535
DB 469 cugucuccaacaauugugagcauccagagagacaguanagccuagaccagagaccgug 528
OY 536 atgtcctcactatcgagaatctgagtcacccctgtgtatgtacaccatcgctcgt 595
DB 529 acguuucacacuaugcggaauugcggaucucugggagugccuaccccuucgucgu 588
OY 596 cagagctactgatacctggttttgacaacatgatagaaccgacactatagtctaagact 655
DB 589 cagaguaucgaaucgccgguuuugacaacaugaugaacaggaagcuguggccacag 648
OY 656 tttctcggttactatgacaggaatgtgccaaggaagacatctcagcgttgagactatg 715
DB 649 uguuucuguuuacauagacaggaauugccagggagagcaugcucgucuggggccaacg 708
OY 716 atccatctactacaacagatctcttccttcctggttcacgactactgttgacagactatgc 775
DB 709 acccgucuaacacacaggauguccugcugugcccgugacagugacagcaguanugc 768
OY 776 aattcactgtgagacagtgctacacacagcgtgtgtgtgtgcatgtaagtgtgacttc 835
DB 769 aguuacacugugagacugucacacacacgugugugugugucugugagugugucuguc 828
OY 836 aagctatcttgataccgctacgctcaagcgtgctggaacgtagcagcaacttccaca 895
DB 829 aggcacauucugagacagcgacacucacacugugugcccgacagcgacgaacucaca 888
OY 896 ttccgcaagctattggagcacaacagaacacgtagcgtgtgattgacatagattgca 955
      : ||||| | | : ||||| ||||| ||||| : ||| : ||||| | | : |||||
```

```

DB 889 uccagcagcgcaauugagccacacagaaaccgaugacgauguuuugacucugcugcga 948
OY 956 accctagctacatgctcactgctgtcttgagatcaacggcaagatgctaccacgaccc 1015
DB 949 accugagcuaacugccacuguguuuugauaauagcaaaugcaaaugacacacgaccc 1008
OY 1016 ccttcgctataccagcagaaggttctgtgacacagtgatctcagaagtgaagacc 1075
DB 1009 ccuucgcuuacacagccaagacagcgcuucuguaaccagugcucacagagugaaauuc 1068
OY 1076 attccagaataatggtatcttgaggatgtgttcattcgtgtgactacaagcgtcttgaca 1135
DB 1069 auuccagaagaaugagucuggggaguuuuaucaggagaguanuuaucgucuuuugaca 1128
OY 1136 gggccaacacctgttggtgtagctaaagcaatctga 1173
DB 1129 gggccaacaacccuggggugcgccaaagccaucuga 1166
```

#### RESULT 5

AAT03006  
ID AAT03006 standard; DNA: 1240 BP.

AAT03006;

13-JUN-1996 (first entry)

Chymosin open reading frame.

al-3; albino mutant; light-regulated; Neurospora; bread mould;  
heterologous gene; expression; control; chymosin; ss.

Mammalian sp.

W09530739-Al.

16-NOV-1995.

09-MAY-1995; 95WO-US05716.

10-MAY-1994; 94US-0240372.

(UYHA-) UNIV HAWAII.

Kato EK, Stuart WD;

WPI; 1995-404108/51.

Nucleic acid for expression of heterologous protein - contains  
albino promoter for light induced expression in filamentous fungi

Example 2; Fig 7; 26pp; English.

The DNA is that of a mammalian gene (open reading frame) encoding  
chymosin. The gene was placed in operable linkage with the al-3  
promoter (see AAT03005). The al-3 gene controls the production of  
geranyl geranyl pyrophosphatase (GGPP) synthetase. GGPP is a precursor  
for carotenoids and xanthophylls. It has been shown that exposure to  
light increases the transcription level of GGPP synthetase 15-45 fold.  
Light activates a number of genes in the common bread mould, Neurospora.  
This can be used to regulate the expression of genes encoding  
heterologous proteins, e.g. chymosin, in recombinant production systems.  
Use of a light-regulated promoter is a simple and effective way to  
control expression and allows timing to be adapted to the physiological  
status of the host.

Sequence 1240 BP; 274 A; 374 C; 339 G; 253 T; 0 other;

Query Match 76.5%; Score 897.2; DB 16; Length 1240;  
Best Local Similarity 87.7%; Pred. No. 7.2e-273;  
Matches 980; Conservative 0; Mismatches 138; Indels 0; Gaps 0;





```
Db 313 gcaaaaaccacccgcttcgaccggaagaagtcgctccacccttcagaaacctggtggcaagc 372
Oy 416 cctgtctatactacagtagcagtagatgcaaggatccttagctatgatccgtca 475
   || ||||| ||||| ||||| ||||| || ||||| ||||| |||||
Db 373 cctgtctatactacagtagcagtagatgcaaggatccttagctatgatccgtca 432
Oy 476 ctgtctcaaatctgtggaatcacaagaacagtaggaacttagcaccgaagaacaggtg 535
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 433 ctgtctcaaatctgtggaatcacaagaacagtaggaacttagcaccgaagaacaggtg 492
Oy 536 atgtctcaaatctgtggaatcacaagaacagtaggaacttagcaccgaagaacaggtg 595
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 493 acgtctcaaatctgtggaatcacaagaacagtaggaacttagcaccgaagaacaggtg 552
Oy 596 cagatctacatgacacgtgttctacacatgatgatcagcagcactagatcgaagact 655
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 553 cagatctacatgacacgtgttctacacatgatgatcagcagcactagatcgaagact 612
Oy 656 tgtctcgtttacatgagagaagaatgccaagagacatgctcaagcttggagatcttg 715
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 613 tgtctcgtttacatgagagaagaatgccaagagacatgctcaagcttggagatcttg 672
Oy 716 atccatctactatacagagatctcttctactgtgttccagatcgtgacagatcgtgac 775
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 673 accgtctactatacagagatctcttctactgtgttccagatcgtgacagatcgtgac 732
Oy 776 aattcactgtgagacagtcacacacacacacacagcagtcgtgtgtgtgtgtgtgtgtgt 835
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 733 agttcactgtgagacagtcacacacacacacacagcagtcgtgtgtgtgtgtgtgtgtgt 792
Oy 836 aagctatcttggataccggtacgttccaaagctgtgtcgaagctagcagcagacatctcaaca 895
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 793 aggcacatcttggataccggtacgttccaaagctgtgtcgaagctagcagcagacatctcaaca 852
Oy 896 tttagcagcagcagtcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 955
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 853 tccagcagcagcagtcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 912
Oy 956 acctagctacatgctacacagtcgttcttggatcaacagcagaagatgtaccactgaccc 1015
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 913 acctagctacatgctacacagtcgttcttggatcaacagcagaagatgtaccactgaccc 972
Oy 1016 cctcgcctatacagcagcagatcaagggtctcgcacagtcgagttccagagtgagaac 1075
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 973 cctcgcctatacagcagcagatcaagggtctcgcacagtcgagttccagagtgagaac 1032
Oy 1076 attccagaagaatgagtccttggagagatgtgtcattcgtggaattacacagcgtctttgaca 1135
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1033 attccagaagaatgagtccttggagagatgtgtcattcgtggaattacacagcgtctttgaca 1092
Oy 1136 ggcccaacaacatcgttggagtagtaagaatctga 1173
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1093 ggcccaacaacatcgttggagtagtaagaatctga 1130

RESULT 8
AAAN20043
ID AAAN20043 standard; DNA; 1460 BP.
XX
AC AAAN20043;
XX
XX 16-DEC-1992 (first entry)
XX
XX DE Pre-prorennin-A gene DNA sequence.
XX
XX KW Pre-pro-rennin; rennin; prorennin; enzyme; EC-3.4.23.4; chymosin;
XX
XX KM protease; milk-clotting enzyme; ss.
XX
XX OS Bos taurus.
XX
XX FT Key Location/Qualifiers
   CDS 205..1350
   FT /*Tag= a
```

```
FT FT /label=
   FT /note= "pre-prorennin-A gene"
   XX GB2091271-A.
   XX
   XX 28-JUL-1982.
   PD
   XX 15-JAN-1982; 82GB-0001120.
   PF
   XX 01-DEC-1981; 81US-0325481.
   PR 16-JAN-1981; 81US-0225717.
   XX
   XX (COLB ) COLLABORATIVE RES INC.
   XX
   XX Alford BL, Mao J, Moir DT;
   XX
   XX WPI: 1982-62028E/30 (62028E).
   DR P-PSDB; AAP20038.
   DR
   XX Transformed cells producing rennin and its precursors - contg.
   PT appropriate recombinant DNA material
   PS
   XX
   XX Bases 1-204 and 1351 to 1460 are attached to the pre-prorennin but
   CC can be removed and are not essential to use of the gene in
   CC expression. The gene may be ligated into plasmid pCDE21 and
   CC expressed in E. coli. The resulting expressed enzyme is a well
   CC known milk-clotting enzyme used in cheese-making.
   XX
   SO Sequence 1460 BP; 329 A; 440 C; 397 G; 294 T; 0 other;

Query Match 76.4%; Score 895.6; DB 3; Length 1460;
Best Local Similarity 87.6%; Pred. No. 2.5e-272;
Matches 979; Conservative 0; Mismatches 139; Indels 0; Gaps 0;

Oy 56 tcgtgtgtgttactcagcgtctgagatcaccgcgactcctctacaagaagtagtctc 115
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 233 tcttgcctctcccgagcgctgagatccacagatccctctgtacaagaagcagttctc 292
Oy 116 tccgtaaggcgctgaagaacatggaactcttagaagaactcttgtagaagaacagatga 175
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 293 tgaagaaggcgctgaagaacatggaactcttagaagaactcttgtagaagaacagatga 352
Oy 176 gcatcagcagaagtagctccggtctcggttgaagtgtcagtcgtgacctaccactacc 235
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 353 gcatcagcagaagtagctccggtctcggttgaagtgtcagtcgtgacctaccactacc 412
Oy 236 ttgatatcgaactcttggagaagatctaccctcogaaccccgctcaagaagttcacgctc 295
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 413 tggatagtcagtagcttggagaagatctaccctcogaaccccgctcaagaagttcacgctc 472
Oy 296 tcttggatagtcgttccctctgactcttggttccctctatctactacgcaagaagtagtct 355
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 473 tgttgacacatggtccctctgactcttggttccctctatctactacgcaagaagtagtct 532
Oy 356 gcaagaaccacacagatctgacccgagagaagtcgltccacacttccagaacttaggcaaac 415
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 533 gcaaaaccacacagatctgacccgagagaagtcgltccacacttccagaacttaggcaaac 592
Oy 416 cctgtctatacactagtgatgagtagcagtagcaagaatcttaggtctatgatccgtca 475
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 593 cctgtctatacactagtgatgagtagcagtagcaagaatcttaggtctatgatccgtca 652
Oy 476 ctgtctcaaatctgtggaatcacaagaacagtaggaacttagcaccgaagaacaggtg 535
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 653 ctgtctcaaatctgtggaatcacaagaacagtaggaacttagcaccgaagaacaggtg 712
Oy 536 atgtctcaaatctgtggaatcacaagaacagtaggaacttagcaccgaagaacaggtg 595
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 713 acgtctcaaatctgtggaatcacaagaacagtaggaacttagcaccgaagaacaggtg 772
```

```
QY 566 cagagtaactcgatgacccgtgtgtttgacaacatgtaaccgacactagtaactgact 655
D 773 cagaagtactcgatgacccgtgtgtttgacaacatgtaaccgacactggtggtcccaagacc 832
QY 656 tgttctggtttactgtagcaggaatggtgcaggaagacatgctcagctgtgagactgatt 715
D 833 tgttctggtttactgtagcaggaatggtgcaggaagacatgctcagctgtgagactgatt 892
QY 716 atccatctactacagatctctctggtgttcacgtgctgacgtgctgacgtgctgacgt 775
D 893 accgctctactacagatggtgttcacgtgctgacgtgctgacgtgctgacgtgctgacgt 952
QY 776 aattcactgtgtagcaggtgttcacacatcagcgtgtgtgtgtgtgtgtgtgtgtgtgt 835
D 953 agttcactgtgtagcaggtgttcacacatcagcgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1012
QY 836 aagctactgtgtagcaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 895
D 1013 aggcacatctgtgtagcaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1072
QY 896 ttccagcaagctgtgtagcaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 955
D 1073 tccagcaagctgtgtagcaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1132
QY 956 accttagctactgtagcaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1015
D 1133 acctgagctactgtagcaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1192
QY 1016 ccttcgctactacagcagcaggaatggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1075
D 1193 ccttcgctactacagcagcaggaatggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1252
QY 1076 attcccaagaatgtagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1135
D 1253 attcccaagaatgtagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1312
QY 1136 gggcacaacaactcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1173
D 1313 gggcacaacaactcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1350
```

```
RESULT 9
AAZ06463
ID AAZ06463 standard; DNA; 2733 BP.
XX
AC AAZ06463:
XX
DT 29-NOV-1999 (first entry)
XX
DE 2.7 Kbp HindIII fragment of pSBSOTPMT (oleosin-chymosin fusion gene).
XX
KW oil-body; lipid body; oleosome; spherosome; separation;
KW fusion protein; heterologous polypeptide; commercial production;
KW plasmid; ds.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT CDS 850..2729
FT /*tag= a
FT /product= "oleosin-spacer-Met-prochymosin"
FT /note= "CDS contains an intron"
FT 850..1202
FT /number= 1
FT 1203..1442
FT /number= 1
FT 1443..2733
FT /number= 2
XX
XX US5948682-A.
XX
XX 07-SEP-1999.
```

```
PF 25-APR-1997; 97US-0846021.
XX
PR 25-APR-1997; 97US-0846021.
PR 22-FEB-1991; 91US-0659835.
PR 16-NOV-1993; 93US-0142418.
PR 30-DEC-1994; 94US-0366783.
XX
PA (SEMB-) SEMBIOSYS GENETICS INC.
PI
PI MoJoney MM:
XX
XX WPI; 1999-517960/43.
XX P-PSDB; AAY33830.
DR
DR Expression of a heterologous polypeptide on an oil body protein is
XX useful for the production of e.g. enzymes, antibodies, hormones
XX
XX Claim 15; Fig 6; 48pp; English.
XX
XX This is the nucleotide sequence of a HindIII fragment containing the
XX oleosin-spacer-Met-prochymosin sequence. This HindIII fragment was
XX joined to a nopaline synthase terminator and cloned into binary vector
XX PCGN1559. The resulting plasmid was called pSBSOTPMT and introduced
XX into A. tumefaciens. The resulting bacterial strain was used to transform
XX B. napus plants.
XX The DNA which encodes a chimeric fusion protein that consists of the
XX oil-body targeting sequence, a transcription regulation sequence and the
XX DNA of the protein of interest can be used to produce antibodies,
XX glycanases, hormones, proteases, protease inhibitors, seed storage
XX proteins, thrombin inhibitors, hirudin, interleukins, chymosin, cystatin,
XX xylanase, carp growth hormone, zein or a collagenase.
XX The enzyme may be cleaved from the oil body protein or used in
XX association with the oil body fraction.
XX Allows production of commercially important proteins on a superior scale
XX to production by conventional systems. The expressed heterologous
XX protein can be easily separated from host cell components
XX due to the use of the oil body as a carrier protein.
XX
XX Sequence 2733 BP; 699 A; 681 C; 684 G; 669 T; 0 other:
```

```
Query Match 76.3%; Score 894.8; DB 20; Length 2733;
Best Local Similarity 88.4%; Pred. No. 6.3e-272;
Matches 971; Conservative 0; Mismatches 127; Indels 0; Gaps 0;
QY 76 gctgagatccgccgactctctcttcaaaaggtaagttcttcgtaaggcgtgaaggaa 135
D 1630 gctgagatccgccgactctctcttcaaaaggtaagttcttcgtaaggcgtgaaggaa 1689
QY 136 catgactctctgaagactctcttgagaacaaacagatgataagcagaagctacc 195
D 1690 catgactctctgaagactctcttgagaacaaacagatgataagcagaagctacc 1749
QY 196 ggtctcggtgaagttgtagcgtgtgacctaccactacactgataagtaacttggg 255
D 1750 ggtctcggtgaagttgtagcgtgtgacctaccactacactgataagtaacttggg 1809
QY 1750 ggtctcggtgaagttgtagcgtgtgacctaccactacactgataagtaacttggg 1809
QY 256 aagatctactcggaaaccccgctcaagaagttacacgttctcttggatactggttctct 315
D 1810 aagatctactcggaaaccccgctcaagaagttacacgttctcttggatactggttctct 1869
QY 316 gactctggtgtccctctatctactgcaagaacatgctgtgaagaacaccaaagattc 375
D 1870 gactctggtgtccctctatctactgcaagaacatgctgtgaagaacaccaaagattc 1929
QY 376 gatccgagaagttgttcaacttccagaaacttagcaaaccttgttataatactagcgt 435
D 1930 gatccgagaagttgttcaacttccagaaacttagcaaaccttgttataatactagcgt 1989
QY 436 acaggtagcatgcaagaatcttaggtatgatactacacgtcactgttccaaacttggac 495
D 1990 acaggtagcatgcaagaatcttaggtatgatactacacgtcactgttccaaacttggac 2049
```

QY	496	atctaaacagacagtagtagactctaaaccaccaagaacacagtgtagtcttactatacgcagaa	555
Db	2050	atccacagacagacagtagagcctctgaacaccacagagagcccgaggagacgtcttcaactatgcgaa	2109
QY	556	ttcgatgtagcctctgtgtatgtagcatccatcgcctgcggctcagagtagctgcgatactctg	615
Db	2110	ttcgacgggagaccctggggagtaggcctccaccctcgcctcgcctcagagtagctgcgatacgcgg	2169
QY	616	tttagacaacatgataaaccagacacactagtagtagctcaagaatgtttctcggtttacatgac	675
Db	2170	tttagacaacatgataaaccagagcaccctcggcggcccaagacgtttctcggtttacatgac	2229
QY	676	aggaaatgcccagagagatgaatgatctcaacgcttgtagatctatgatcatcctactacaagaa	735
Db	2230	aggaaatgcccagagagatgaatgatctcaacgcttgtagatctatgatcatcctactacaagag	2289
QY	736	tctcttaacttggtgttccagatctgtgcagacgagtagtagcaattcaactgtgtagacgttgc	795
Db	2290	tccctgcacttggtgttccagatctgtgcagacgagtagtagcaattcaactgtgtagacgttgc	2349
QY	796	accatacagcgtgtgtaggtgtgtgtatcgtatgaaggtgtagtgcacagctatactcttgataccggt	855
Db	2350	accatacagcgtgtgtaggtgtgtgtatcgtatgaaggtgtagtgcacagctatactcttgataccggtc	2409
QY	856	acgtcccaagctgtgtcggacctaagcagcgcacatctcacaatccagcaagctatgtgagcc	915
Db	2410	acgtcccaagctgtgtcggacctaagcagcgcacatctcacaatccagcaagctatgtgagcc	2469
QY	916	acacagaacacagtagcgtgtgtaggtttgacatagatttgagacaacctgtagctacgtgctaca	975
Db	2470	acacagaacacagtagcgtgtgtaggtttgacatagatttgagacaacctgtagctacgtgctaca	2529
QY	976	gttgtctttgagatcaacacggcagaagatgataccacatcacccctcgcctatacagcagcag	1035
Db	2530	gttgtctttgagatcaacacggcagaagatgataccacatcacccctcgcctatacagcagcaca	2589
QY	1036	gatacaagggtctgtcaccaagtgtagttccagagtgagaaaccatccccaagaatgtagtcttg	1095
Db	2590	gatacaagggtctgtcaccaagtgtagttccagagtgagaaatccatccccaagaatgtagtcttg	2649
QY	1096	ggagatgtgtttcatctcgttgtagtactaacagcgtctttgacaaggccaacaacctgtgtgg	1155
Db	2650	ggagatgtgtttcatctcagagagatatacagcgtctttgacaaggccaacaacctgtgtggg	2709
QY	1156	ctagctaaagaacatctga	1173
Db	2710	ctagctaaagaacatctga	2727
RESULT 10			
AA014051			
ID	AA014051 standard; DNA; 1210 BP.		
XX	AA014051;		
AC	06-JAN-1992 (first entry)		
XX	Rennin gene.		
DE	Prorennin; alpha-SI-casein gene; insulin-like growth factor I; IGF-I;		
XX	mammary gland; ss.		
KW	Synthetic.		
OS	EPA51823-A.		
XX	16-OCT-1991.		
PN	10-APR-1991;		
XX	91EP-0105702.		
PD	19-APR-1990;		
XX	90DE-4012526.		
PR	11-APR-1990;		
XX	90DE-4011751.		

[illegible]



QY	DB	Sequence	Score	DB ID	Length	Match	Local Similarity	Conservative	Mismatches	Indels	Gaps
QY	894	tcacgacgagccatcgtggaacacacagcaaccacgtaacgctgagcttgacatgacgacga	953								
QY	956	acctagatctacatgcctacagcttgctctcttgatgatacaccgcaagatgtaccactgacc	1015								
DB	954	acctagatctacatgcctacgcttgctctcttgatgatacgaataatgtaccactgacc	1013								
QY	1016	cctcgcctataccgacgagatcaaggctctcgtacacagtgatccagatgagtgaaacc	1075								
DB	1014	cctcgcctataccgacgacgaacgaaggctctcgtacacagtgatccagatgagtgaaacc	1073								
QY	1076	attccacgaatgtgactcttgagagatgtgttcattcgtgagatcactacagcgtctcttgaca	1135								
DB	1074	attccacgaatgtgactcctctggggagatgttttcattcgcgagatcactacagcgtctcttgaca	1133								
QY	1136	gggcacaacacctcgtctggcgtgactgaagcaatcttga	1173								
DB	1134	gggcacaacacctcgtctggcgtgactgaagcaatcttga	1171								
RESULT 12											
AAAN91188	ID	AAAN91188 standard; DNA; 2727 BP.									
XX	AC	AAAN91188;									
XX	DT	15-JUN-1990 (first entry)									
XX	DE	BamHI/SalI insert of Kluyveromyces plasmid PAB309.									
XX	KW	Kluyveromyces; PDM100PC; chymosin; tissue plasminogen activator.									
XX	OS	Kluyveromyces lactis.									
XX	FT	Key	Location/Qualifiers								
XX	FT	CDS	409..1781								
XX	FT	/*tag= a									
XX	PN	EP301670-A.									
XX	PD	01-JAN-1989.									
XX	PF	28-JUL-1988; 88EP-0201632.									
XX	PR	28-JUL-1987; 87US-0078539.									
XX	PA	(KONN ) GIST-BROCADES NV.									
XX	PI	van den Berg JA, van Ooyen AJJ, Rietveld K;									
XX	DR	WPI; 1989-033565/05.									
XX	DR	P-PSDB; AAP94376.									
XX	PT	Kluyveromyces host cells for producing polypeptide(s) -									
XX	PT	used for highly efficient prodn. of eg chymosin tissue									
XX	PS	plasminogen activator or human serum albumin.									
XX	PS	Disclosure; 56pp; English.									
XX	CC	BamHI/SalI insert, incorporated into plasmids PAB309 with									
XX	CC	promoters, terminators and a G418 resistance marker fused to a ADHI									
XX	CC	promoter from S.cerevisiae.									
XX	CC	The Kluyveromyces expression systems provide highly efficient secretion									
XX	CC	and processing of a wide variety of proteins.									
XX	CC	Sequences identical to those published in EP301669.									
XX	XX	Sequence 2727 BP; 752 A; 654 C; 572 G; 749 T; 0 other;									
QY	57	ctttcctgtatcaacgctcgtctgaagataccacccatctcctctcaacaaagttaadgtctct	116								

[illegible]





KW microbial vector; ss.  
XX Bos taurus.  
OS  
XX Key Location/Qualifiers  
FT CDS 1..1140  
FT /\*tag= a  
XX BE897201-A.  
XX PD 03-NOV-1983.  
XX 30-JUN-1983; 83BE-0017731.  
XX 13-APR-1983; 83US-0484539.  
XX 01-JUL-1982; 82US-0394433.  
XX (GENE-) GENEX CORP.  
XX WPI; 1983-820813/47.  
XX P-PSDB; AAP30013.  
XX Isolated chymosin or rennin and prochymosin genes - plasmid(s)  
PT which replicate in prokaryotic organisms, esp. *Escherichia coli*,  
PT and organisms used for chymosin biosynthesis  
XX Claim 7; Page 33-36; 43pp; French.  
XX The inventors claim isolated chymosin (rennin) and prochymosin genes  
CC from calves, and plasmids contg. the genes which are capable of  
CC replicating in a prokaryotic organism. The prokaryotic organism is  
CC pref. an *Escherichia* species, esp. *E. coli* p Gx 1225 (NRRL B-15061).  
CC The microorganisms transformed by the plasmid are also claimed.  
SQ Sequence 1289 BP; 310 A; 391 C; 327 G; 260 T; 1 other;  
Query Match 75.9%; Score 889.8; DB 4; Length 1289;  
Best Local Similarity 87.2%; Pred. No. 1.6e-270;  
Matches 975; Conservative 0; Mismatches 143; Indels 0; Gaps 0;  
QY 56 tctgttcgttactacgtctgtgagatcacccgcattctctcttcaagaagtaattctc 115  
DB 23 tcttcgtctctccagggcgtcgagatcacagagatccctctgtacaagaagcttc 82  
QY 116 tccgttaaggcgtcgaagacatctgactctagaagactctctgcagaacaacagttatg 175  
DB 83 tgaagaaaggcgtcgtgagacatctggtctctgagagacttctctgcagaacaacagttatg 142  
QY 176 gcatacgaagaagtaactcgcgttcgtgtaagtgtgtagcgtgtgcacttaccactacc 235  
DB 143 gcatcagcagcaagtaactcgcgttcgtgtaagtgtgtagcgtgtgcacttaccactacc 202  
QY 236 ttgatagtaaatattcttggaagatctactctggaaccccgctcaagaagttcaccttc 295  
DB 203 tagataagtaactcttgggaagatctactctggaaccccgctcaagaagttcaccttc 262  
QY 296 tctttgatactgttctcctgactctgtggttccctctactactgcagaagcaatgcct 355  
DB 263 tgtttgacactgtcctctgactctgttacttctgtaacctactactgcagaagcaatgcct 322  
QY 356 gcaagaaccaccaagaatcgatccgagaagaagtcgtccactccagaacttaggcaaac 415  
DB 323 gcaaaaaccaccagcgtctcgaccggagaagtcgtccactccagaacttaggcaaacg 382  
QY 416 ccttgtctatacactcaggtacaggtagcatgcaagaatcttagctatgataacgtca 475  
DB 383 ccttgtctatacactcaggtacaggtagcatgcaagaatcttagctatgataacgtca 442  
QY 476 ctgtctccaacattgtggacatccaacagatagagacttagagccaagaacagcagttg 535  
DB 443 ctgtctccaacattgtggacatccaacagatagagacttagagccaagaacagcagttg 502

QY 536 atgtcttaccatcagcaagaattcgatcgatccttctgtatgagatcacccatcgtcgt 595  
DB 503 agcttccactatgcngaatttcgcgcggatccctggggatgagcctacccttcgtcgtc 562  
QY 596 cagagtactcgataccctgtgtttacaacaatgatgatgaacccagactagtagtaagaact 655  
DB 563 cagagtactcgataccctgtgtttacaacaatgatgatgaacccagactagtagtaagaact 622  
QY 656 tgttcctgttatactgacagaagaatgagcagaagacatgctacactgtgaactatgt 715  
DB 623 tgttcctgttatactgacagaagaatgagcagaagacatgctacactgtgaactatgt 682  
QY 716 atccatctactacaagaagatccttcaactgtgttcacgtacactgtgacagtaacttgc 775  
DB 683 accgcctctactacaagaagatccttcaactgtgttcacgtacactgtgacagtaacttgc 742  
QY 776 aattcactgtgacagttcaccatcagcaggtgtgtgttctgtacatgtgaagttgagttc 835  
DB 743 agttcactgtgacagttcaccatcagcaggtgtgtgttctgtacatgtgaagttgagttc 802  
QY 836 aagctatcttgataccggtacgttccaaagctgttcgaacctcagcagcagatcttcaaca 895  
DB 803 aggcacatctgacacagcagcactccaaagctgtgtcggccagcagcagcagcactcaaca 862  
QY 896 ttcaagaagctattgtgagccacacagaaccagtaacgtgagttgacatagattgcgaca 955  
DB 863 tccagaagcctattgtgagccacacagaaccagtaacgtgagttgacatagattgcgaca 922  
QY 956 accctagctatactgactacagttgtcttctgtgagatcaacggcaagatgtaaccaactgacc 1015  
DB 923 accctagctatactgactacagttgtcttctgtgagatcaacggcaagatgtaaccaactgacc 982  
QY 1016 cctcgcctataccagcagaagatcaagggtctctgaccagtgagatccagagttgagaac 1075  
DB 983 cctcgcctataccagcagaagatcaagggtctctgaccagtgagatccagagttgagaac 1042  
QY 1076 attccagaagaatggtcttggagatgtgttcattcgttgagtaactacagcgtcttgaca 1135  
DB 1043 attccagaagaatggtcttggagatgtgttcattcgttgagtaactacagcgtcttgaca 1102  
QY 1136 gggccaacaacactcgttggctagctaaagcaatctga 1173  
DB 1103 gggccaacaacactcgttggctagctaaagcaatctga 1140  
RESULT 15  
AA004683  
ID AA004683 standard; DNA; 1098 BP.  
XX AA004683;  
AC  
XX 05-OCT-1990 (first entry)  
DT  
XX  
DE Sequence encoding calf pro-rennin.  
KW Pro-rennin; ds.  
XX  
XX Bos taurus.  
OS  
XX JP02109984-A.  
PN  
XX 23-APR-1990.  
PD  
XX 01-JAN-1988; 88JP-0302176.  
PF  
XX 01-JAN-1988; 88JP-0302176.  
PR  
XX (BEPP/) BEPPU T.  
PA  
XX WPI; 1990-168358/22.  
DR P-PSDB; AAR05080.  
XX  
XX Complex plasmid and microbe - contains calf pro-rennin cDNA.  
PT

XX Disclosure; 32; 13bp; Japanese.  
 XX  
 CC Protein product may be expressed in E.coli expression system from  
 CC plasmid pBR322.  
 XX  
 SQ Sequence 1098 BP; 252 A; 327 C; 300 G; 219 T; 0 other;

Query Match 75.6%; Score 886.8; DB:11; Length 1098;  
 Best Local Similarity 88.0%; Pred. No. 1.3e-269;  
 Matches 966; Conservative 0; Mismatches 132; Indels 0; Gaps 0;

Qy 76 gctagatcacccgcatctctctacaaagtaagctcccgtaagcgctaaagaa 135  
 |||||  
 Db 1 gctagatcacccgcatctctctacaaagtaagctcccgtaagcgctaaagaa 60  
 Qy 136 catggaactctagaagactcttcgaagaacacagatgagcatcagcagaactacc 195  
 |||||  
 Db 61 catggaactctagaagactctctcgaagaacacagatgagcatcagcagaactacc 120  
 Qy 196 ggcttcggtgaagtgctgaagtgctcacttaccactacactctgatatgcaacttggg 255  
 |||||  
 Db 121 ggcttcggtgaagtgctgaagtgctcacttaccactacactctgatatgcaacttggg 180  
 Qy 256 aagatctaccctcggaaccccgctcaagaagctcaccgtctcttgaatcgtgtccctc 315  
 |||||  
 Db 181 aagatctaccctcggaaccccgctcaagaagctcaccgtctcttgaatcgtgtccctc 240  
 Qy 316 gactcttggttccctctatctactgcaagaacatgctcctgcaagaacacacaaagatlc 375  
 |||||  
 Db 241 gactcttggttccctctatctactgcaagaacatgctcctgcaagaacacacaaagatlc 300  
 Qy 376 gatccgagaaagtcgtccactctccaaacttaaggcaaacccctgtctatacactaagtc 435  
 |||||  
 Db 301 gatccgagaaagtcgtccactctccaaacttaaggcaaacccctgtctatacactaagtc 360  
 Qy 436 aacagtgacatgacgaacatcttagctatgatacgttcaactgtctccaaacttgttgac 495  
 |||||  
 Db 361 aacagtgacatgacgaacatcttagctatgatacgttcaactgtctccaaacttgttgac 420  
 Qy 496 attcaacagacagtagaacttagcaaccaagaacacagtgatgtcttcaactatgacaa 555  
 |||||  
 Db 421 attcaacagacagtagaacttagcaaccaagaacacagtgatgtcttcaactatgacaa 480  
 Qy 556 ttgcattgcatccttggtatgataccatcgctcgctcagatgatacctgtg 615  
 |||||  
 Db 481 ttgcattgcatccttggtatgataccatcgctcgctcagatgatacctgtg 540  
 Qy 616 ttgacaacatgatacgaacgaacacttagctagctcaagaactgtctcggtttacatgac 675  
 |||||  
 Db 541 ttgacaacatgatacgaacgaacacttagctagctcaagaacactgtctcggtttacatgac 600  
 Qy 676 aggaatgccaagagacatgctcagcttgaagctatgttccatcctactacacaga 735  
 |||||  
 Db 601 aggaatgccaagagacatgctcagcttgaagctatgttccatcctactacacaga 660  
 Qy 736 tctcttactgggttcacagctcgtcagcagatcgtgcaatcctgttgaagctgtc 795  
 |||||  
 Db 661 tctcttactgggttcacagctcgtcagcagatcgtgcaatcctgttgaagctgtc 720  
 Qy 796 acaatcagcggt 855  
 |||||  
 Db 721 acaatcagcggt 780  
 Qy 856 acgttcaagctgt 915  
 |||||  
 Db 781 acgttcaagctgt 840  
 Qy 916 acacagaacacagtagcgt 975  
 |||||  
 Db 841 acacagaacacagtagcgt 900

Qy 976 gtgtcttggatcacaacgcaagatgtaacccactgacccctccgctatataccagccag 1035  
 |||||  
 Db 901 gtgtcttggatcacaacgcaagatgtaacccactgacccctccgctatataccagccag 960  
 Qy 1036 gatcaagggtctcgcaccagtgatgatacagagtgagaacacatcccaagaatgtgacttg 1095  
 |||||  
 Db 961 gatcaagggtctcgcaccagtgatgatacagagtgagaacacatcccaagaatgtgacttg 1020  
 Qy 1096 ggaagatgttcatctcgtgagtaactacagcgcttcttgcaaggccaagaactcgttggg 1155  
 |||||  
 Db 1021 ggaagatgttcatctcgtgagtaactacagcgcttcttgcaaggccaagaactcgttggg 1080  
 Qy 1156 ctgactaaagcaatctga 1173  
 |||||  
 Db 1081 ctgactaaagcaatctga 1098

Search completed: July 31, 2002, 21:11:57  
 Job time: 25748 sec

...

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 11:56:19 ; Search time 5848.1 Seconds  
(without alignments)  
2707.192 Million cell updates/sec

Title: US-09-643-755B-1  
Perfect score: 1173  
Sequence: 1 atgaacttcttaagctttt.....ggctagctaaagaactctga 1173

Scoring table:  
IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*  
1: em\_estbda:\*  
2: em\_esthum:\*  
3: em\_estln:\*  
4: em\_estlnu:\*  
5: em\_estor:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hlc:\*  
9: gb\_estl:\*  
10: gb\_est2:\*  
11: gb\_hlc:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	387.2	33.0	472	10	BG938320
2	358	30.5	631	9	AI892188
3	354.8	30.2	430	10	BG937697
4	337.2	28.7	727	9	AI326975
5	323.8	27.6	399	10	BG937723
6	300.8	25.6	1347	11	AK004109
7	300.6	22.3	383	10	BG938086
8	262	22.3	479	9	AA028632
9	249.6	21.3	505	9	AI479358
10	246	21.0	546	9	AI324867
11	242.4	20.7	517	9	AI385490
12	236.2	20.1	1385	11	AK008959
13	236	19.4	619	9	AW868692
14	227.4	19.3	1388	11	AK008886
15	226.8	18.6	414	9	AI322423
16	217.8	17.6	636	9	AM868716
17	206.4	17.6	814	10	BI761345

18	205.8	17.5	583	10	BE841742
19	204.4	17.4	910	10	BF299798
20	202.8	17.3	622	10	BF373831
21	201	17.1	716	9	AM012992
22	199.4	17.0	603	10	BE841740
23	199	17.0	702	9	AM867433
24	198.6	16.9	551	10	BE841714
25	196.4	16.7	547	10	BM069200
26	196.4	16.7	548	10	BM053683
27	190.6	16.2	391	9	AA969042
28	190	16.2	433	10	W34201
29	190	16.2	713	10	BF121716
30	187.4	16.0	422	10	W10274
31	185	15.8	666	10	BE841761
32	182.6	15.6	594	9	AM863768
33	182.2	15.5	578	10	BE840800
34	182	15.5	564	9	AM868711
35	181	15.4	736	10	BI821121
36	180.8	15.4	847	10	BI490133
37	180.8	15.4	616	10	BE841758
38	180.4	15.4	547	10	BE841657
39	179.6	15.3	576	9	AM868673
40	179.6	15.3	960	10	BI517784
41	179.2	15.3	751	9	AI076336
42	178.8	15.2	815	10	BI818940
43	176.4	15.0	666	9	AM959943
44	176.2	15.0	847	10	BI762154
45	176	15.0	572	10	BF373812

#### ALIGNMENTS

##### RESULT 1

BG938320 472 bp mRNA linear EST 11-JUN-2001  
LOCUS IAB015E12 Bovine Abomasum cDNA Library Bos taurus cDNA 5', mRNA  
DEFINITION sequence.

ACCESSION BG938320

VERSION BG938320.1

KEYWORDS GI:14337692

SOURCE EST.

ORGANISM COW.

Bos taurus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;

Bovidae; Bovinae; Bos.

REFERENCE 1 (bases 1 to 472)

AUTHORS Moore,S.S., Hansen,C., Li,C., Fu,A., Meng,Y. and Li,G.

TITLE CDNA's from bovine abomasum tissue

JOURNAL Unpublished (2001)

COMMENT Contact: Dr. Stephen Moore

. Beef Genomics Laboratory

Dept of AFNS, University of Alberta

410 Agri/For, Dept of AFNS, U of A, Edmonton, AB, T6G 2P5, Canada

Tel: 780 492 0169

Fax: 780 492 4265

Email: smoores@afns.ualberta.ca

The sequence best matches gb:BOCHYMOA (bovine chymosin a (fennin)

trna) in main database at high score of 928.0 and E-value of 0.0

PCR Primers

FORWARD: M13 Forward

BACKWARD: M13 Reverse

Seq primer: T3 primer

High quality sequence stop: 472

POLYA-NO.

Location/Qualifiers

1..472

/organism="Bos taurus"

/db\_xref="taxon:9913"

/clone\_lib="Bovine Abomasum cDNA Library"

/sex="Two males and one female mixed"

/tissue\_type="Gastrointestinal tissue (GIT)"

/cell\_type="Epithelial"



DB	541	AAATGGGCGGTTCACAGTGGACAGGATCAATCAATGTAAGTGAAGTGGCTTTCACAG	600
QY	828	tgatggtcaagctatcttggataccggtac	857
Db	601	TGGCTGCCCTGCCGTTCTGGACACAGGAC	630
RESULT	3		
LOCUS	BC937697	430 bp	mRNA
DEFINITION	LA005E01	Bovine Abomasum cDNA Library	Bos taurus cDNA 5', mRNA
ACCESSION	BC937697		
VERSION	BC937697.1	GI:14337069	
KEYWORDS	EST.		
SOURCE	COV.		
ORGANISM	Bos taurus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovinae; Bos.		
AUTHORS	1 (bases 1 to 430)		
TITLE	Moore,S.S., Hansen,C., Li,C., Fu,A., Meng,Y. and Li,G.		
JOURNAL	CDNA's from bovine abomasum tissue		
COMMENT	Unpublished (2001)		
CONTACT	Contact: Dr. Stephen Moore		
DEPT	Beef Genomics Laboratory		
AFNS	Dept of AFNS, University of Alberta		
AGRI/FOOT	410 Agri/Foot, Dept of AFNS, U of A, Edmonton, AB, T6G 2P5, Canada		
TEL	Tel: 780 492 0169		
FAX	Fax: 780 492 4265		
EMAIL	Email: smoores@afns.ualberta.ca		
SEQUENCE	The sequence best matches gb:BOVCHYMOA (bovine chymosin a (rennin) mRNA) in main database at high score of 844.0 and E-value of 0.0		
PCR	PCR Primers		
FORWARD	FORWARD: M13 Forward		
BACKWARD	BACKWARD: M13 Reverse		
SEG	Seg primer: T3 primer		
QUALITY	High quality sequence stop: 430		
POLYA	POLYA-No.		
FEATURES			
SOURCE	location/Qualifiers		
	1..430		
	/organism="Bos taurus"		
	/db_xref="taxon:9913"		
	/clone_lib="Bovine Abomasum cDNA Library"		
	/sex="Two males and one female mixed"		
	/tissue_type="Gastrointestinal tissue (GIT)"		
	/cell_type="Epithelial"		
	/dev_stage="Young adult"		
	/lab_host="XLI-BlueMF"-strain		
	/note="Organ: Abomasum; Vector: Uni-22APYR; Site_1: EcoR I; Site_2: Xho I"		
BASE COUNT	97 a 130 c 119 g 84 t		
ORIGIN			
Query Match	30.2%	Score 354.8;	DB 10; Length 430;
Best Local Similarity	89.1%;	Pred. No. 3.5e-90;	
Matches 383;	Conservative 0;	Mismatches 47;	Indels 0; Gaps 0;
QY	605	cgatccgcttcttgcacacatgtagaacgacacactgtgagctaaagactgttctcg	664
Db	1	CGATACCCGCTGTTACACATGATGAACAGGACCTGGTGCCCAAGACCTGTTCTCGG	60
QY	665	ttatcatggaagagatgagcagagagagatgcatgacgcttgagactatgtacatct	724
Db	61	TTTACATGGACAGGAATGGCCAGGAGACATGCTCAGCGTGGGGCCATGACCCGTCCT	120
QY	725	actacacagatcttctactggttccagatcactgtgacagcagatcgtgcaatcactg	784
Db	121	ACTACACAGGGGCTCTGCATGCTGGCCCGTGCACAGTGCACGACGATGAGCTTCACAG	180
QY	785	tgagacaggtacacatcagcggtgtggtgtgtgtgcatgtggaagtgtgagatcattct	844

OY	181	TGGACAGTGTACCAATCAGCGGTTGTGTGTGGCTGTGAGGGTGCGTGTACAGGCATTC	240
OY	845	tggataccggtatcgtccaaagctgtgcgaacctagacagacattctcaacattcagaag	904
Dd	241	TGGACACGGGACCTCCAAAGCTGTGTGGGCCACAGACGACATCTCCATCACATCCAGAG	300
OY	905	ctatggagggccacaagaaccagtgcgtgtgatatttgacatatgttcgacaacacttagt	964
Dd	301	CCATTGGAGCCACACAGAACCAGTACGATGATTTTGACATCCACTGCCAACCTGAGCT	360
OY	965	acatctaccacagttgctctttgataccaacggaagatgccaccatgaccccctcgct	1024
Dd	361	ACATCCCCCATCTGTGCTCTTTGATGATCATGTGCATAAATGTACCCACTGACCCCCTCGCT	420
OY	1025	ataccagcca	1034
Dd	421	ATACCAGCCA	430
RESULT	4		
LOCUS	A1326975/c		
DEFINITION	A1326975	727 bp mRNA linear EST 23-DEC-1998	
ACCESSION	m31e03.x1	Soares mouse p3MNF19.5 Mus musculus cDNA clone	
VERSION	IMGE:482524	3' similar to TR:Q28075 Q28075 CHROMOSIN C.; mRNA sequence.	
KEYWORDS	A1326975		
SOURCE	A1326975.1	GI:4061404	
ORGANISM	house mouse.		
REFERENCE	Mus musculus		
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. I (bases 1 to 727)		
	Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubugue,T., Geisler,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theisinger,B., Wylie,T., Lennon,G., Soares,B., Willson,R. and Waterston,R.		
TITLE	The WashU-HMI Mouse EST Project		
JOURNAL	Unpublished (1996)		
COMMENT	Contact: Marra M/Mouse EST Project WashU-HMI Mouse EST Project Washington University School of MedicineP 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@wustl.edu This clone is available royalty-free through LNL; contact the IMGE Consortium (info@imge.llnl.gov) for further information. MGI:293268 This clone was previously sequenced on the 5' end only, this new data is from the 3' end Possible reversed clone: similarity on wrong strand High quality sequence stop: 446. Location/Qualifiers 1..727 /organism="Mus musculus" /db_xref="taxon:10090" /clone="IMAGE:482524" /dev_stage="19.5 dpc total fetus" /clone_lib="Soares mouse p3MNF19.5" /note="Vector: pRT33 (pharmacia) with a modified polylinker. Site.1: Not I - oligo(dn) primer 15', TGTACCAATCTGAAGTGGGCGCGCGCATTTTTTTTTTTTTTTT 3', double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pRT33 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M.Palima Bonaldo. RNA was kindly provided by Dr. Minoru Ko (Wayne State University)."		







```

Db      700 AAGGCACGATGCTGATGCTAGTGGAGGTGATTCCTCCTACTACATGAGAGGTTTAC   759
          ||| ||||| ||| || | ||| ||||| ||||| ||| |||||
Qy      745 tgggtccagtcactctgycagcagtaactgcaatcaccttgcagtgcatccaacatcasc   804
          ||| ||||| ||| ||| ||| ||||| ||||| ||| |||||
Db      760 TGGGTACCAGTGCTCCAAAGGCCAGCTACTGCGCAATTAGCTGTGATAGCATCTCCATGAAT   819
          ||| ||||| ||| ||| ||| ||||| ||||| ||| |||||
Qy      805 ggtgttgtttgtgatctatgtgaagtgtgtagtatgaactatcttgataccggtfagcgacaag   864
          ||| ||||| ||| ||| ||| ||||| ||||| ||| |||||
Db      820 GGGGAGGTCATTTGCTGTGATGTGGCTGCCAAGGTATTATGACACAGGAGACCTCTTG   879
          ||| ||||| ||| ||| ||| ||||| ||||| ||| |||||
Qy      865 ctgttcgacctagaacgagcattctcaacatlcaagcagcatattgagccacacagaaac   924
          ||| ||||| ||| ||| ||| ||||| ||||| ||| |||||
Db      880 CTGACCGGCCCCCGAAGCTCCATCGTTAACATCCAGATCTAATTGTGTGCCAAGGCTTCT   939
          ||| ||||| ||| ||| ||| ||||| ||||| ||| |||||
Qy      925 cagtacggtgtgagtttgcatagatattgcgacaaccttgatcatactacagttgcttct   984
          ||| ||||| ||| ||| ||| ||||| ||||| ||| |||||
Db      940 GGTGACGGCGAGTACTTCTCTCAAGTGTGACACATCAACACCTGCTGTGATTTGTCTTC   999
          ||| ||||| ||| ||| ||| ||||| ||||| ||| |||||
Qy      985 gagatcaacgycagaagtatgtaccacactgaccccccttcgcctatatccaggccagatcaaagg   1044
          ||| ||||| ||| ||| ||| ||||| ||||| ||| |||||
Db      1000 ACCATTCCGCACTGTATCTTACCCTCCAGTGCACCACTGCTCATCTCGAAGAAGATGGTCA   1059
          ||| ||||| ||| ||| ||| ||||| ||||| ||| |||||
Qy      1045 ttctgcaccagttgattccagagttagaac-----cattcccagaatcgg   1089
          ||| ||||| ||| ||| ||| ||||| ||||| ||| |||||
Db      1060 CACAATTCACAGAGCACACTTTGAGGAGGGCATGATGATCCATCAGACCTTGATGATGTGG   1119
          ||| ||||| ||| ||| ||| ||||| ||||| ||| |||||
Qy      1090 atcttggagagatgtgttcatctgtgtgtagtactacagcgtcttgcagagggccacaacctc   1149
          ||| ||||| ||| ||| ||| ||||| ||||| ||| |||||
Db      1120 GTGCTGGGGGATGTCTTCTCTGAGCGCTGATTTTACACCGTGTGTGATCGGGCAATAAACAAG   1179
          ||| ||||| ||| ||| ||| ||||| ||||| ||| |||||
Qy      1150 gttaggcctaagcattcga   1173
          ||| ||||| ||| ||| ||| ||||| ||||| ||| |||||
Db      1180 ATTGCTGTGCTCTGCTGCATGA   1203
          ||| ||||| ||| ||| ||| ||||| ||||| ||| |||||

RESULT    7
LOCUS     BG938086                               383 bp   mRNA   linear   EST_11-JUN-2001
DEFINITION LAbOIIIA08 Bovine Abomasum cDNA Library Bos taurus cDNA 5', mRNA sequence.
ACCESSION BG938086
VERSION   BG938086.1 GI:14337458
KEYWORDS  EST.
SOURCE    cow.
ORGANISM  Bos taurus
           Eumariota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
           Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 383)
AUTHORS  Moore,S.S., Hansen,C., Li,C., Fu,A., Meng,Y. and Li,G.
TITLE    cDNA's from bovine abomasum tissue unpublished (2001)
JOURNAL   Contact: Dr. Stephen Moore
COMMENT   . Beef Genomics Laboratory
           Dept of AFNS, University of Alberta
           410 Agri/FOR, Dept of AFNS, U of A, Edmonton, AB, T6G 2P5, Canada
           Tel: 780 492 0169
           Fax: 780 492 4265
           Email: smoores@afns.ualberta.ca
           The sequence best matches gb:BPJ9786 (Bos primigenius procyomisin
           mRNA, complete cds) In main database at high score of 735.0 and
           E-value of 0.0
PCR PRIMERS
FORWARD: M13 Forward
BACKWARD: M13 Reverse
Seq primer: T3 primer
High quality sequence stop: 383
POLYA=No.
Location/Qualifiers
1..383
/organism="Bos taurus"
/db_xref="taxon:9913"

```

Query Match	25.6%;	Score 300.6;	DB 10;	Length 383;
Best Local Similarity	88.3%;	Pred. No. 9.2e-75;		
Matches 338;	Conservative 0;	Mismatches 44;	Indels 1;	Gaps 1;
QY 599	agtaactgataccctgtgtgtttgaacaacatgatgatgaacgacaccctagtagcttaagacttgt			
Db 1	AGTACTGATACCGCGTGTGTGACAACATGATGAAACAGGACGACCTGGTGGCCCAAGACCTGT			
QY 659	tctcggttcaatgtagacaggaatggccagggagagatgtctaogctttggaactttgtc			
Db 61	TCTCGGTTTACATGAGACAGGAATGGCCGAGAGCATCTCAAGTGGGGGCCATCGAC			
QY 719	catctactacaagaagactctctactagtgttcacagtcacgtgtgcagagtaactgtcaat			
Db 121	CGTCTTACTACACAGGGTCTCTCGCACTGGGTGCCGTATACAGTGCAGCAGTACTGGCAGT			
QY 779	tcaactgtgagacagtgtaaccacatcagcaggtgtgtgtgtgtcatgtlgaagggtgagtc			
Db 181	TCACGTGTGAGACAGTGTACACATCACCGGTGTGTGTGTGTGGCTGTGAGGGTGGCTGTAGG			
QY 839	ctatcttgatataccggtatcgtccaaagctgtgtccagcctcaggagacattctcaacctc			
Db 241	CCATCCTTGAGACAGGGGACCTCCAACTGTGGTGGGCCCAAGAGGCACATCTCAACATCC			
QY 899	agcaagctatgtagagcaacacagaacacagtaacgtgtgtgtgtacat-agattgcagacaac			
Db 301	AGCAGGCCATTTGGAGCCACACAGAACCAAGTACGGTGAATTTGACATCTGACATGGCAGAAC			
QY 958	cttagctacatgctctacagttgt 980			
Db 361	CTGAGCTACATGCCACATCTGTGT 383			
RESULT 8				
AA028632	479 bp	mRNA	linear	EST 11-SEP-1996
LOCUS	ml14c08.1 Soares mouse p3MwF19.5 Mus musculus cDNA clone			
DEFINITION	IMAGE:463102.5' similar to SW:CHM_BOVIN P00794 PROCHYMOSIN A AND B			
PRECUSORS	;; mRNA sequence.			
ACCESSION	AA028632			
VERSION	AA028632.1			
KEYWORDS	EST.			
SOURCE	house mouse.			
ORGANISM	Mus musculus			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
	1. (bases 1 to 479)			
	Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,			
	Giesel,S., Kucaba,T., Lacy,M., Le-M., Martin,J., Morris,M.,			
	Schellensberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,			
	Theisinger,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and			
	Waterston,R.			
	The WashU-HHMI Mouse EST Project			
	Unpublished (1996)			
TITLE	WashU-HHMI Mouse EST Project			
JOURNAL	Washington University School of Medicine			
COMMENT	4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108			
	Te1: 314 286 1800			
	Fax: 314 286 1810			
	Email: mouseest@watson.wustl.edu			



```

RESULT 10
AI324867/c
LOCUS
DEFINITION
AI324867 546 bp mRNA linear EST 23-DEC-1998
m182901.x1 Soares mouse p3NMF19.5 Mus musculus cDNA clone
IMAGE:473065 3' similar to TR:Q28075 Q28075 CHYMOSIN C.; mRNA
sequence.
ACCESSION
AI324867 GI:4059296
VERSION
KEYWORDS
SOURCE
ORGANISM
Mus musculus.
house mouse.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 546)
REFERENCE
Marrin,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
The Mashu-HHMI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
Mashu-HHMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:283809
This clone was previously sequenced on the 5' end only, this new
data is from the 3' end
High quality sequence stop: 451.
FEATURES
Source
Location/Qualifiers
1..546
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:473065"
/clone_lib="Soares mouse p3NMF19.5"
/dev_stage="19.5 dpc total fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Vector: pT73D (Pharmacia) with a modified
polylinker; Site_1: Not I - oligo(dt) primer 15',
TGTTCACATCGAATGAGGAGCGCGCCGATTTTCTTTTCTTTT 3',
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M.Fatima Bonaldi. RNA was kindly provided by
Dr. Minoru Ko (Mayne State University)."
```

```

BASE COUNT
134 a 140 c 140 g 132 t
ORIGIN
Query Match 21.0%; Score 246; DB 9; Length 546;
Best Local Similarity 70.9%; Pred. No. 4,1e-59;
Matches 343; Conservative 0; Mismatches 135; Indels 6; Gaps 1;

OY 690 gagcagtcacagctgagctatgatccatctactacacagatctcttcaactgggt 749
|||||
DB 546 GAGCATGCTCACACATGGGGGCGATCGATCATCTTCAATAGGCTCACTGCATGGGT 487
|||||
OY 750 tccagtcactgtgcagcagctactgcaattcaactgtgtgacagtgatcaccatgaagcggt 809
|||||
DB 486 GCCTGTGACAGTACAGGAGGATATTGACAGTTTACAGAGTGCACAGATCAATCATGTGA 427
|||||
OY 810 gattgttgatgtgagtgagtgatgtcaagctatcttgatgacggtgacgagtgatgt 869
|||||
DB 426 AGTGTGCTGTGTCAGAGGTGGCTCCCTGCTTGTGACACAGCAATGCTCCTGTTGAC 367
|||||

FEATURES
source
1..517
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:482524"
/clone_lib="Soares mouse p3NMF19.5"
/dev_stage="19.5 dpc total fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Vector: pT73D (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dt) primer 15'
High quality sequence stop: 442.
location/Qualifiers
1..517
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:482524"
/clone_lib="Soares mouse p3NMF19.5"
/dev_stage="19.5 dpc total fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Vector: pT73D (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dt) primer 15'

OY 870 cgaacctagcagcgacatcttcaacattcagcaagctatiggaagccaacagcagta 929
|||||
DB 366 GGGGCTGTGTGAGACATCTCCATATTCACAGATGATGAGACTGTGCAAGCCATTA 307
|||||
OY 930 cgggtgattgacatagatgtcgacaaccttagtatacatgctacagttgtcttgatg 989
|||||
DB 306 TGACCACTTTACATTCGACTCTGTGAGGCTGTGACATCATGCGCCACCGCTGCTTGA 247
|||||
OY 990 caacgcaagatgtaaccactgaacccctccgcctataccagcagatgaaggtctcg 1049
|||||
DB 246 CCATGTAGGAGATTCCTCCACTGCACCTACGCTTATACCAACCAAGTACAGGCTCTG 187
|||||
OY 1050 caccagtgatccagagtggaacacatcccaagaatgagatcttgagagatgtgtcat 1109
|||||
DB 186 CTCCAGTGGCTTC-----AAGCAGGCTCCACATGTGGATCTTGAGGATGTTTCAT 133
|||||
OY 1110 tctgtgactacacagcgctcttgacaagggcaacaacctgtgtggtagtaagaacat 1169
|||||
DB 132 TCGGAGTTATACAGTGTGTTGACAGAGCCACATGCTGTGGCTGGCCAGGCTAT 73
|||||
OY 1170 ctga 1173
|||||
DB 72 CTGA 69

RESULT 11
AI385490 517 bp mRNA linear EST 27-JAN-1999
LOCUS
DEFINITION
m181e03.y1 Soares mouse p3NMF19.5 Mus musculus cDNA clone
IMAGE:482524 5' similar to TR:Q28950 Q28950 PREPROCHYMOSIN
PRECURSOR.; mRNA sequence.
ACCESSION
AI385490
VERSION
AI385490.1 GI:4198953
KEYWORDS
SOURCE
EST.
house mouse.
ORGANISM
Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 517)
REFERENCE
Marrin,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Rittler
,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
Waterston,R. and Wilson,R.
The Mashu-NCI Mouse EST Project 1999
Unpublished (1999)
Contact: Marra M/Mashu-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:293268
This read is a RESEQUENCE of a previously sequenced mouse clone
This read has been verified (found to hit its original self in the
correct orientation)
Possible reoriented clone: similarity on wrong strand
Seq primer: -40RP from Gibco
High quality sequence stop: 442.
location/Qualifiers
1..517
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:482524"
/clone_lib="Soares mouse p3NMF19.5"
/dev_stage="19.5 dpc total fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Vector: pT73D (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dt) primer 15'
```



```

CDs      66..1244
         /note="data source:SPRR, source key:P20142, evidence:ISS
         homolog to GASTRISIN PRECURSOR (EC 3.4.23.3) (PEPSINOGEN
         c)
         putative"
         /codon_start=1
         /protein_id="BA025990.1"
         /db_xref="gi:12843461"
         /translation="MKMNVVALICLPLEALIRVPLKMKRSIRETMKEGVLDKEL
         NHRYPDQGHFGKPGDYSLVPEPMAYMAYSGEISIGTPONFVLIEDTSGSNLMY
         SSVYCOACTTHTRNPNKSSRYTQGGFSLQYGGSLTGFXYDLTREVOSIOVN
         OERGSLENPERGTRFYVAQPDGIMGLAYPGISGGCATALOGMDEGALSQLFGVYG
         SQQSGNGOIVREGVDENLTGELTWIPTYOELWQITDDELIGMDSKSSGSCQ
         GIYDTGSLVMPAOYLINELLOTIGAEGEYGYFVSCDSVSLPLTFLVNGVQFPL
         SPSSYIIQEGSCMGLIESILNANESGQPLMIIGDVFLRSYVAFPMGNRVLAPSV
         "
polyA_signal 1369..1374
              /note="putative"
polyA_site    1385
              /note="putative"
BASE COUNT   281 a 441 c 336 g 327 t
ORIGIN

```

```

Query Match      20.18; Score 236.2; DB 11; Length 1385;
Best Local Similarity 53.7%; Pred. No. 4,2e-56;
Matches 602; Conservative 0; Mismatches 483; Indels 36; Gaps 4;

```

```

QY 76 gctgagatcaccgcatctctctctacaaagtaagtcctccgtaagggcgtgaagaa 135
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 111 GCACCTTGTGATCAGGGGCCCTCCCTGAGAAATGAAGATATCCGGAGACCATGAAGAA 170

QY 136 catggaacttagaagaactcttgcaagaacaagatagagcatcagcagaagtaactcc 195
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 171 CAAGGTGTCTCAAGAACTTCTCTGAAGAACCAAGATGATGACCTGGCCAGAAATACCAAC 230

QY 196 ggccttgatgaagtgtctagcgtctcacttaccac-----ctacctgatatgtaa 246
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 231 TTGTGGCAAGTTTGGTGTGATCAGTGTATGAGGCCCATGGCCTATATGATGATCTTCC 290

QY 247 tacttggagaagatctactcctggaaaccccgctcaagaagtaactcagttctcttatact 306
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 291 TACTATGTTGATGATCAGATCGGAGATCCACCCAGAACTTCTGCTCTTTTGCACACT 350

QY 307 ggttcctctgactctgggtctccctctactactgcaagaagcaatgacctgcaagaacac 366
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 351 GGCCTCTCAACACTGTGGGTGTCTTCTGTACTGCCAGAGGAGGCTGCACACACACAC 410

QY 367 caaagatctgactcggagaagtcgtccacttccagaacttaaggcaaaccttgcctata 426
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 411 ACCGCTACACCCCGACAGATCTCTCCACTACTCAAGGGCAGACCTTTCCTCG 470

QY 427 cactttagatagatgtagatgcaagaactctagactatgataccgtactcgttcccaac 486
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 471 CAGTACGGCACCGGACGCTTACCGGCTTCTTCGGCTATGACACTCTGAGAGTCCAAAGC 530

QY 487 attgtagacatccaacagacagtagaacttagcaccacaagaacagatgattcttccacc 546
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 531 ATCCAGGTCCTTAACACAGAGAGTTGGCCTGAGTAGAATGAGCTGGACCACTATTTGTC 590

QY 547 tatgagaatctgactgcatccttggtatgcatatcccatcgctcgctcaagatcactcg 606
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 591 TACGCCCAATTTGACGGGATCATGGGCTTGCCCTACCCCGGCTCTTCTTCAAGGGGGCGCC 650

QY 607 atacttggtttgacaatgatgtaacccagacccagtagctcaagacttgcttcggtt 666
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 651 ACCACGGCTTGCACAGGATGTTGGGGAGGGGCGCTGTGTCCACGCCCTCTTCGGTGTTC 710

QY 667 tacatgagacaagaatgagccagagagcat-----gtcacactggaagcatatgaccca 720
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 711 TACCTTGGCAGCGACAGAGGGGTCTAAAGCGGGGAGATGTGTGTGGGTGGCGTGACGAG 770

QY 721 tctactacaagaatctcttcaactggtttccagtaactgtgacagcagtaactgccaatcc 780

```

```

DB 771 AACCTGATACCTGGCGAGCTCACCCTGATTCCTCTCTACCCAGAGACTTTTACTGGCAGATC 830
QY 781 actgtgacagatgtccaccatcagcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 834
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 831 ACCATTGACGACCTTCTTATTTGGCAACGACGCTTGTGCTGTCTCTCTCTCTCTCTCTCTCT 890
QY 835 caagctctcttgataccggttagcttcaagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 894
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 891 CAAGGCAATTTGATGACACAGGACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 950
QY 895 attagaagaagctatgtgagccacacagacaagcagatgagatgtgtgtgtgtgtgtgtgtgt 954
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 951 CTTTTCGACAGACCATAGAGAGCCACAGAGAGAGATATGAGACATTTTGTGTAGCTGGGAC 1010
QY 955 aacctagctactatgctctacagatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1014
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1011 AGCGTCAGTACCTGCTGCTACCCCTCAGTTTGTCTCTCAATGATGTGTCTCCCTCTCTCA 1070
QY 1015 ccttcgccttataccagcagcagatcagaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1067
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1071 CCTCTTCTCTCATCATCATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1130
QY 1068 -----tgaagaccattcccaagaatgagctctgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1119
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1131 CTGACAGCTGAGAGTGGCGACAGCCCTCTGATCTCTCGGGAGATGCTTCTCAGGCTTAC 1190
QY 1120 tacagcgtctttagacaagggccaacaacctcgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1160
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1191 TATGCTGTCTTGACATGAGCAATATACAGGCTGGGCTTTC 1231

```

```

RESULT 13
AM868692      619 bp      mRNA      linear      EST 22-MAY-2000
LOCUS      MRL-SN0064-150500-004-a03 SN0064 Homo sapiens cDNA, mRNA sequence.
DEFINITION      AM868692
ACCESSION      AM868692.1 GI:8002744
VERSION      EST.
KEYWORDS      human.
SOURCE      Homo sapiens
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

```

```

REFERENCE
1 (bases 1 to 619)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bata,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.

```

```

TITLE      Shotgun sequencing of the human transcriptome with ORF expressed
      sequence tags
JOURNAL      Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE      20202663
COMMENT      Contact: Simpson A.J.G.
      Laboratory of Cancer Genetics
      Ludwig Institute for Cancer Research
      Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
      Brazil
      Tel: +55-11-2704922
      Fax: +55-11-2707001
      Email: asimpson@ludwig.org.br
      This sequence was derived from the FAPESP/LICR Human Cancer Genome
      project. This entry can be seen in the following URL
      (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2-MRL-SN0064-150
      500-004-a03&t3=2000-05-15&t4=1)
      Seq primer: puc 18 forward
      High quality sequence start: 10
      High quality sequence stop: 619.
      Location/Qualifiers
        1..619
        /organism="Homo sapiens"
        /db_xref="taxon:9606"

```

```

FEATURES
      source

```



```

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="MGI:1895666"
/db_xref="taxon:10090"
/clone="2210410L06"
/sex="male"
/tissue_type="stomach"
/clone_lib="Riken full-length enriched mouse cDNA library"
/dev_stage="adult"
66..1256
/note="data source:SPRR, source key:P20142, evidence:ISS
homolog to GASTRUCSIN PRECURSOR (EC 3.4.23.3) (PEPSINOGEN
C)
putative"
/codon_start=1
/protein_id="BAB25952.1"
/db_xref="GI:12843350"
/translation="MKMMVALLCLPLEALLRVPPKKKSTRETKKEGVAKDPLK
NKRIDFGKTHGKFGDYSLVEPMATMDASTYGEISIGTPONFLVLPFGSSNLMV
SSVYCOSEACTTHRYNPSSSTYTQGFSLQYGTSLTGFEGDVLTVOS1QVFN
OEGLSENEPTNMFVAOEDGIMGLAYPGISGGATTAAGMGEALSOPLGVYLG
SOOGSNGOIVFGEVDENLYTGLTWTWIPYQELWOTIIDFLIGNASGSSGCO
GIYDTGTSILVMPAOYLINELLOTIGAOGEGYGVFVSCDSVSLPTLFVINGVDFPL
SPRSYIIQEBGSCMGLIESLNAESQPLMIGDVPFLRSYVASTWATIGMALPLIS
RDI"
polyA_signal
1368..1373
/note="putative"
polyA_site
1388
/note="putative"
BASE COUNT      282 a      441 c      337 g      328 t
ORIGIN

```

```

Query Match      19.4% Score 227.4; DB 11; Length 1388;
Best Local Similarity 53.9%; Pred. No. 1.3e-53;
Matches 546; Conservative 0; Mismatches 446; Indels 21; Gaps 3;

QY      76 gctgagatcacccgcatctctctacaaagtaagtcctcgtlaagcgcgtgaagaa 135
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB     111 GCAGCTTTGATCAGGTCGCCCGCGAAGAAATGAAGATATCCGGAGACCATGAGAA 170

QY     136 catgagactctagaagaactcttcgagaagaacaacagatgcatcagagaagttactcc 195
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB     171 CAAAGTGTCTCAAAAGACTTTGTGAAGAACCAAGATATGACCTCGGCAGAAATACAC 230

QY     196 ggcctggtgaagtcgtcagcgtgccaacttaacca-----ctaacctgatatcaa 246
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB     231 TTTGGCAAGTTGGTGAATACATGATCTATATAGCCCATAGGCTATATGATGCTTCC 290

QY     247 tacttgggaagatctacactcggaacccgcctcaagaagttcacgcttctatctgatact 306
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB     291 TACTATGATGATACATGACATCGGAGACTCCACCCCAAACTTCGTCCTTTGACACT 350

QY     307 gtttcctcgaactctggttccctctatctactgaagcaagcaatgctcgtgaagaacac 366
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB     351 GGGCTCTCCACACCTGTGGGTCTTCTGTCTACTGCGAGAGGAGGCTTGACACACACAC 410

QY     367 caaagatcgatccgagaagtcgccaacttcagaacttaggcaaaccttctctata 426
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB     411 ACCGGTACAAACCCAGCAATCTCTCCACTATACACTCAAGGGCAGACCTTCTCCCTG 470

QY     427 caactagttacaggtacgatacgaagaactcttaggctatagctacgctcgtctccaac 486
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB     471 CAGTACGGCAGCCGAGCCTTACCGGCTTCTCGGCTATGACACTCTGAGAGTCCAAAGC 530

QY     487 attgtgaacattcaagaagcagtaggacttagcaccacaagaccagtgatgtcttcacc 546
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB     531 ATCCAGGTCCCTAACCAGGAGTTGGCTGAGTAGAATGAGCCTGGCACCACCAATTTTGC 590

QY     547 tatcagaattcgaatgcactcctgtagtgcataccatcgctgcgctcagagttactcg 606
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB     591 TAGCCCAATTATGACGGGATATGGGCTGGGCTACCCGGGCTGTCTTACAGGGGGGCC 650

```

```

QY     607 atacctgtgtttgacaacatgataaccgacacacttagtgcataagaactgtctcggtt 666
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB     651 ACCACCCCTTGACAGGGCATGTTGGGGAGAGGCTGTCGTGCCAGCCCTTCGTGATG 710

QY     667 tacttggacag-----gaatgagcagagagacatgtgtcaagcttgtagcattgtatcca 720
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB     711 TACCTTGGCAGCCAGCAGGGGCTTAACGGCGGGCAGATGTGTTGGGTGGCTGACAG 770

QY     721 tccactacacagaagatctctcaactgtggttcaactcgtgtcagcagtagtgcgaattc 780
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB     771 AACCTGTACATGCGAGAGCTACCTGGATTCCTTCACACCGAGAGCTTTACTGCGACATC 830

QY     781 actgtggaacagtgatccatcagcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 834
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB     831 ACCATTTGCGACTTCTTATTTGGCAACCGACCTCTGCTGCTGCTCTCTCTGCTGCTG 890

QY     835 caaactctcttgatctacggtgagtcgaagctgtgtgagcagtagcagcagatctcacc 894
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB     891 CAAAGCATTTGACACAGGACGACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 950

QY     895 attcaagaagctattggaagccacacagaacagcagtagcaggttgcataagattgcgac 954
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB     951 CTTCGTGAGACCATAGAGAGCCCGAGAGAGATGAGACATTTTGTGACCTGCGAC 1010

QY     955 aacctagctacatgcctctacagtgcttctgtgagatacaacggaagatgtaccacgtacc 1014
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB     1011 AGCGTCAGTACCTGCTTACCTACCTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCA 1070

QY     1015 cctccgcgtataccagcagcagatcaagggtcttcgacacagtgatccagag 1067
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB     1071 CCGTCTCTACATCATCAGAGAGAGGCTCTGACATGCTGGGTCTGAGAG 1123

```

```

RESULT 15
AI322423      414 bp      mRNA      linear      EST 23-DEC-1998
LOCUS      m182b01.y1 Soares mouse p3NMPL1.5 Mus musculus cDNA clone
DEFINITION IMAGE:473065 5' similar to TR:Q28075 Q28075 CHYMOSIN C.; mRNA
sequence.
ACCESSION   AI322423
VERSION     AI322423.1 GI:4056852
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus.
REFERENCE   1 (bases 1 to 414)
AUTHORS    Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wyllie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
            The WashU-HMI Mouse EST Project
            Unpublished (1996)
            Contact: Marra M/Mouse EST Project
            WashU-HMI Mouse EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@wustl.edu
            This clone is available royalty-free through LNL; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:283809
            This read is a RESSEQUENCE of a previously sequenced mouse clone
            This read has been verified (found to hit its original self in the
            correct orientation)
            putative full length read
            vector to vector length is 655
            Seq primer: -40RP from Gibco
            High quality sequence stop: 395.
FEATURES
Location/Qualifiers
I..414

```









PA (SEMB-) SEMBIOSYS GENETICS INC.  
XX  
PI Van Rooijen G, Keon RG, Boothe J, Shen Y;  
XX  
XX WPI: 2001-226621/23.  
DR N-PSDB; AAS00569, AAS00570.  
XX  
XX Producing chymosin in seeds of plants such as rice, flax, rape seed, by  
PT transforming plant cell with a nucleic acid encoding chymosin operably  
PT linked to transcription regulator and terminator sequences -  
XX  
XX Example 1; Fig 1; 56pp; English.  
XX  
XX The sequence represents a bovine chymosin polypeptide. Chymosin can be  
CC produced in a plant seed through introduction of a chimeric nucleic acid  
CC molecule, comprising a nucleic acid sequence encoding a chymosin  
CC polypeptide operatively linked to transcription regulator and terminator  
CC sequences, into a plant cell. The sequences are useful for producing  
CC plant seeds, in particular seeds of soybean, rape seed, sunflower,  
CC cotton, corn, tobacco, alfalfa, wheat, barley, oats, sorghum, Arabidopsis  
CC thaliana, potato, flax/linseed, safflower, oil palm, groundnut, Brazil  
CC nut, coconut, castor, coriander, squash, jojoba and rice.  
XX  
SQ Sequence 390 AA:  
  
Query Match 100.0%; Score 2064; DB 22; Length 390;  
Best Local Similarity 100.0%; Pred. No. 6, 6e-187;  
Matches 390; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 MNETKSPFFVFLFCGQYEVAVTHAAETRIPLLYKSKLRALKKEHGLEDFLOKQOYGI 60  
Db 1 mnetksfpyafllcfgyfvavthaaetrlplykkslralkkehglledflokgqygi 60  
  
QY 61 SSKTSGGEVAVPLTNTLDQYFGKITLYGPPQEFVLPFTGSSDFWVPSIYCKSNACK 120  
Db 61 ssktsggevasvplntnyldsqyfgklygcpqefvlpftgssdfwvpsiycksnack 120  
  
QY 121 NHOFPDRKSTFONLGRPLSIHYGTSGMOGILGYDTVTASNIVDIOQTGVLSTQEPGDV 180  
Db 121 nhofdpkrkstfognlgrplsihygtsgmgilgydvtvasnivdiqvtglstqepgdv 180  
  
QY 181 FTYAEFGILGMATPPLASERISYVFNDMMNRHLVADLFSVYMDRNGQESMLTGAIDP 240  
Db 181 ftyaeifgllgmaypslaseyslvfndmmnrhlvaqdlfsvymdrngqesmltlgaidp 240  
  
QY 241 SYTSGSLHWVPVTVQYQWQFTVDSVTISGVVACGGCOALIDTGSTKLVPSSDIINTIQ 300  
Db 241 sytsgslhwvpvtvqyqwfvtvdsvtisgvvaceggcqaalldtgstklvpsddilniq 300  
  
QY 301 QALGATONQYGEFDIDCNLSYMPVVEIENGKMPPLTPSAVTSODQGFCTSGFQSENHS 360  
Db 301 qalgatnqygefddidcnlsympvvefengkmppltpsaavtsodqgfctsgfqsenshs 360  
  
QY 361 QKWILGDVFTREYTSVFDNRANLVGLAKAI 390  
Db 361 qkwilgdvftreyysvfdnrannlvglakai 390  
  
RESULT 2  
AAP94376 2  
ID AAP94376 standard; protein: 458 AA.  
XX  
XX AAP94376;  
XX  
XX 15-JUN-1990 (first entry)  
XX  
XX BamHI/Sali insert of Kluyveromyces plasmid pAB309.  
XX  
XX Kluyveromyces; pDM100PC; chymosin; tissue plasminogen activator.  
XX  
XX Kluyveromyces lactis.  
XX  
OS

PN EP301670-A.  
XX  
XX 01-JAN-1989.  
PD  
XX  
XX 26-JUL-1988; 88EP-0201632.  
PF  
XX 28-JUL-1987; 87US-0078539.  
PR  
XX 28-JUL-1987; 87US-0078539.  
XX  
XX (KONN ) GIST-BROCADES NV.  
PA  
XX  
XX van den Berg JA, van Ooyen AJJ, Rietveld K;  
PI  
XX  
XX WPI: 1989-033565/05.  
DR  
XX  
XX N-PSDB; AAN91188.  
DR  
XX  
XX Kluyveromyces host cells for producing polypeptide(s) -  
PT used for highly efficient prodn. of eg chymosin tissue  
PT plasminogen activator or human serum albumin.  
XX  
XX Disclosure; : 56pp; English.  
XX  
XX BamHI/Sali insert gene product, incorporated into plasmids pAB309 with  
CC promoters, terminators and a G418 resistance marker fused to a ADHI  
CC promoter from S.cerevisiae.  
CC The Kluyveromyces expression systems provide highly efficient secretion  
CC and processing of a wide variety of proteins.  
CC Sequences identical to those published in EP301669.  
XX  
SQ Sequence 458 AA:  
  
Query Match 93.7%; Score 1933; DB 10; Length 458;  
Best Local Similarity 99.7%; Pred. No. 2, 2e-174;  
Matches 366; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
QY 24 HAAETRIPLLYKSKLRALKKEHGLEDFLOKQOYGISKTSYSGFEVAVPLNTYLDQY 83  
Db 24 haaetrlplykkslralkkehglledflokgqygisisktsygsfgevavplntnyldsqy 151  
  
QY 84 FGKITLYGPPQEFVLPFTGSSDFWVPSIYCKSNACKNHOFPDRKSTFONLGRPLSIH 143  
Db 84 fgkitlygppqefvlpftgssdfwvpsiycksnacknhofdpkrkstfognlgrplsih 211  
  
QY 152 FGKILYGTTPPEFVLIDTGSDFWVPSIYCKSNACKNHQFDPRKSTFONLGRPLSIH 211  
Db 152 fgkilygtppqefvlpftgssdfwvpsiycksnacknhqfdprkstfognlgrplsih 211  
  
QY 144 YGTSGMOGILGYDTVTASNIVDIOQTGVLSTQEPGDVFTYAEFGILGMATPPLASERIS 203  
Db 144 ygtsgmgilgydvtvasnivdiqvtglstqepgdvftyaefgllgmaypslaseysi 271  
  
QY 204 PVFDMMNRHLVADLFSVYMDRNGQESMLTGAIDPSYTYTSGSLHWVPVTVQYQWQFTVD 263  
Db 204 pvfdmmnrhlvaqdlfsvymdrngqesmltlgaidsytsylhwvpvtvqyqwfvtvd 331  
  
QY 264 SVTISGVVACGGCOALIDTGSTKLVPSSDIINTIOAIGATONQYGEFDIDCNLSYM 323  
Db 264 svtisgvvaceggcqaalldtgstklvpsddilniqqaigatnqygefddidcnlsym 391  
  
QY 324 PTVVEIENGKMPPLTPSAVTSODQGFCTSGFQSENHSQKWILGDVFTREYTSVFDNRANL 383  
Db 324 ptvvefengkmppltpsaavtsodqgfctsgfqsenshqkwilgdvftreyysvfdnrannl 451  
  
QY 392 PTVVEIENGKMPPLTPSAVTSODQGFCTSGFQSENHSQKWILGDVFTREYTSVFDNRANL 451  
Db 392 ptvvefengkmppltpsaavtsodqgfctsgfqsenshqkwilgdvftreyysvfdnrannl 451  
  
QY 384 VGLAKAI 390  
Db 452 vglakai 458  
  
RESULT 3  
AAI33830 3  
ID AAI33830 standard; protein: 545 AA.  
XX  
XX AAI33830;  
XX  
XX 29-NOV-1999 (first entry)  
XX  
XX Oleosin-spacer-Met-prochymosin amino acid sequence.  
XX  
DE

XX	oil-body; lipid body; oleosome; spherosome; separation;
KW	fusion protein; heterologous polypeptide; commercial production;
KW	plasmid; oil-body targeting sequence.
XX	
OS	Synthetic.
OS	Arabidopsis thaliana.
XX	
FH	Key
FH	Location/Qualifiers
FT	Protein
FT	/label= Oleosin targeting sequence
FT	119..175
FT	Protein
FT	/label= Oleosin protein
FT	176..181
FT	cleavage_site
FT	/note="Thrombin cleavage site"
FT	182..483
FT	/label= Met-Prochymosin
XX	
PN	US5948682-A.
XX	
PD	07-SEP-1999.
XX	
PF	25-APR-1997; 97US-0846021.
XX	
PR	25-APR-1997; 97US-0846021.
PR	22-FEB-1991; 91US-0659835.
PR	16-NOV-1993; 93US-0142418.
PR	30-DEC-1994; 94US-0366783.
XX	
PA	(SEMB-) SEMBIOSYS GENETICS INC.
XX	
PI	Moloney MM;
DR	WI: 1999-517960/43.
DR	N-PADB; AAZ06463.
PT	Expression of a heterologous polypeptide on an oil body protein is
PT	useful for the production of e.g. enzymes, antibodies, hormones
XX	
PS	Example 10; Fig 6; 48pp; English.
XX	
CC	This is the amino acid sequence of a HindIII fragment containing the
CC	Oleosin-spacer-Met-prochymosin sequence. This HindIII fragment was
CC	joined to a nopaline synthase terminator and cloned into binary vector
CC	PGCM1559. The resulting plasmid was called PSBSORPMT and introduced
CC	into A.tumefaciens. The resulting bacterial strain was used to transform
CC	B.napus plants.
CC	The DNA which encodes a chimeric fusion protein that consists of the
CC	oil-body targeting sequence, a transcription regulation sequence and the
CC	DNA of the protein of interest can be used to produce antibodies,
CC	glycanases, hormones, proteases, protease inhibitors, seed storage
CC	proteins, thrombin inhibitors, hirudin, interleukins, chymosin, cystatin,
CC	xylanase, carp growth hormone, zein or a collagenase.
CC	The enzyme may be cleaved from the oil body protein or used in
CC	association with the oil body fraction.
CC	allows production of commercially important proteins on a superior scale
CC	to production by conventional systems. The expressed heterologous
CC	protein can be easily separated from host cell components
CC	due to the use of the oil body as a carrier protein.
XX	
Sequence	545 AA;
90	

Query Match	93.4%	Score 1928:	DB 20:	Length 545:
Best Local Similarity	97.6%:	Pred. No. 8.3e-174:		
Matches 368:	Conservative 1:	Mismatches 6:	Indels 2:	Gaps 1:
QY	16	GOYFAVAVTHA--AEITRIPLVKGKSLRKALKEHGLEDFLOKQOYGISISKYSGFEVASV	73	
	159			
Db	169	gqhtlvprgsmaelttrplykyskrlrkalkenhgllledtlqgqyiskysgfgfevasv	228	
	169			
QY	74	PLTNVLDQSYRQKLYLGGTPPOEFTVLEDTGSSDFWVPSLYICKSNACKNHQRPDKRSSTF	133	
	74			

Db	229	plnqldsqyfgkilylgpppgeflvldtgssdcfwpsjycksnacknqhrgfoprkestf	288
QY	134	ONLQKPLSIHNGTSSMGICIGEDTVYTSNIYVDIOQTWGLSTOEPBGVFTYAEEDGILGMA	193
Db	289	qnlgprlsibhygtsmgqllgydcctvcvsnlvdlqqvlgstqepgdvflvyaetdgilgma	348
QY	194	YPSLASEYSIVPFPNMNMNRHLVAODLFPSVYMDRNGQESMLTGLADPSYVTGSLHWVPY	253
Db	349	ypslaseysivpfnmnmnrhlvaqdltsvymdrngqesmltqlgaidspsytgslhvwpyt	408
QY	254	VQQTWQGFVDSVTTSISGVYVACEGGCOALIDTGTSKLVGPSDDILNIDQALGATQNGYGER	313
Db	409	vqyqwgfvdsvttsisgvvvaecgcqalldgtgtsklygpsddilnlqqaigatqngygef	468
QY	314	DIDCDNLISYMPVTVEFIEIGKKWPLTPSAVTSQDDGFCFTSGFQSEHNSOKMILGDVFIREY	373
Db	469	dldcdnlisymptvvefmgkmyrltpsaetsyqdqgftcstgfsqeshnsqkmlldgvfirey	528
QY	374	YSVDFRANNLVGLAKAI 390	
Db	529	ysvdftrannlvglakai 545	

Query Match	Score	DB	Length
Best Local Similarity	98.4%;	Pred. No. 1.8e-173;	
Query Match	93.1%;	Score 1922;	DB 5; Length 381;

Matches 364; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 21 AVTHAAETTRIPLYKGSLSRKALKEHGLLEDFLOKQOYGISSKYSGFGEVASVPLTNVLD 80  
::: |||||  
Db 12 alsqgaeltirpilykxslrkalkenhllledfllkqgyisksygfgevasvpltnlyld 71

QY 81 SOYFGKITLTPPOEFVLEFDTGSSDFWVPSIYCKSNACKNHQRPDRKSSTFQNLGKPL 140  
|||  
Db 72 sqyfgkilylgtppqefvlifdgtssdfwvpsiycksnacknhqrfdrkssltfnlgkpl 131

QY 141 SIHGTGSMOGLIGDVTVSNIVDIQOTVGLSTQEPGDVFTVAEFDGILGMAYPSLASE 200  
|||  
Db 132 sihygtsmgqilgydctvsniivdiqvtglstqepgdvftlyaeftgdlgmaypslase 191

QY 201 YSIFVFNMMNRHLVAODLFESVYMDRNGQESMLTGAIDPSYTGSLHWVTVTVOQYQWF 260  
|||  
Db 192 yslpfdmmnrhlvagdflsvymdrngqesmltgaidspsytcslhwvvtvqyqwf 251

QY 261 TVDSVTISGVVAVCEGCGQALIDTGTSLKLVGPSSDILNIQAIATONQYGEFIDICDNL 320  
|||  
Db 252 tvdsvtisgvvavcegcqalldtgtsklvgpsddlningaigtatqygefdiddcni 311

QY 321 SYMPTVFEINGKMPPLTPSAVTSODGFCFSGFQSENHSQKWLIGDVFIREYVSVPDRA 380  
|||  
Db 312 symptvfeingkmpltpsaytsdqgfcfsgfsehsqkwlilgdvfireyvsvidra 371

QY 381 NNLVGLAKAI 390  
|||||  
Db 372 nnlvglakai 381

RESULT 5  
AAP40078  
ID AAP40078 standard; Protein: 375 AA.

XX AAP40078:

XX 02-FEB-1992 (first entry)

XX Sequence encoded by prochymosin gene.

XX Prochymosin expression vector; E.coli trp operon; chymosin.

XX Bos taurus.

XX EP121775-A.

XX 17-OCT-1984.

XX 07-MAR-1984; 84EP-0102451.

XX 09-MAR-1983; 83JP-0038439.

XX (BEPP/) BEPPU T.

XX Beppu T, Uozumi T, Nishimori K, Shimizu N, Kawaguchi Y;

XX Hidaka M;

XX WPI: 1984-258001/42.

XX N-PSDB; AAN40055.

XX Expression plasmid comprising prochymosin gene and vector -

XX useful for transforming Escherichia coli for prochymosin prodn.

XX PS Disclosure; Fig 1; 59pp; English.

XX The inventors claim the prochymosin gene comprising a nucleotide  
CC sequence from (a) the 1st codon (GCT) to the 365th codon (ATC), or  
CC (b) the 5th codon (CGG) to the 365th codon (ATC); and recombinant  
CC plasmids harboured by Escherichia coli strains deposited as FERB BP-  
CC 262, -263 and -264. Any portion of the nucleotide sequence as  
CC described in AAN40055 can be used. Also claimed is a vector derived  
CC from plasmid pBR322. Typically plasmid pCR501 is obtd. from pOCT 2.

CC The transcriptional direction of pOCT 3 is opposite to that of  
CC pOCT 2; it is clockwise in pOCT 2 (5' to 3') whereas counter  
CC clockwise in pOCT 3.

XX Sequence 375 AA;

Query Match 93.1%; Score 1921; DB 5; Length 375;  
Best Local Similarity 98.4%; Pred. No. 2,2e-173;  
Matches 364; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 21 AVTHAAETTRIPLYKGSLSRKALKEHGLLEDFLOKQOYGISSKYSGFGEVASVPLTNVLD 80  
::: |||||  
Db 6 alsqgaeltirpilykxslrkalkenhllledfllkqgyisksygfgevasvpltnlyld 65

QY 81 SOYFGKITLTPPOEFVLEFDTGSSDFWVPSIYCKSNACKNHQRPDRKSSTFQNLGKPL 140  
|||  
Db 66 sqyfgkilylgtppqefvlifdgtssdfwvpsiycksnacknhqrfdrkssltfnlgkpl 125

QY 141 SIHGTGSMOGLIGDVTVSNIVDIQOTVGLSTQEPGDVFTVAEFDGILGMAYPSLASE 200  
|||  
Db 126 sihygtsmgqilgydctvsniivdiqvtglstqepgdvftlyaeftgdlgmaypslase 185

QY 201 YSIFVFNMMNRHLVAODLFESVYMDRNGQESMLTGAIDPSYTGSLHWVTVTVOQYQWF 260  
|||  
Db 186 yslpfdmmnrhlvagdflsvymdrngqesmltgaidspsytcslhwvvtvqyqwf 245

QY 261 TVDSVTISGVVAVCEGCGQALIDTGTSLKLVGPSSDILNIQAIATONQYGEFIDICDNL 320  
|||  
Db 246 tvdsvtisgvvavcegcqalldtgtsklvgpsddlningaigtatqygefdiddcni 305

QY 321 SYMPTVFEINGKMPPLTPSAVTSODGFCFSGFQSENHSQKWLIGDVFIREYVSVPDRA 380  
|||  
Db 306 symptvfeingkmpltpsaytsdqgfcfsgfsehsqkwlilgdvfireyvsvidra 365

QY 381 NNLVGLAKAI 390  
|||||  
Db 366 nnlvglakai 375

RESULT 6  
AAP40218  
ID AAP40218 standard; Protein: 381 AA.

XX AAP40218:

XX 25-JAN-1992 (first entry).

XX Sequence of rennin encoded by recombinant CGF4.

XX Yeast expression vector; GAL1 promoter; Saccharomyces cerevisiae.

XX Bos taurus.

XX GB2137208-A.

XX 03-OCT-1984.

XX 28-FEB-1984; 84GB-0405129.

XX 28-FEB-1983; 83US-0470911.

XX (COLB ) COLLABORATIVE RES INC.

XX Botstein D, Davis RW, Fink GR, Taunton-Rigby A, Knowlton RG;

XX Mao JT, Moir DT, Golf CG;

XX WPI: 1984-245517/40.

XX N-PSDB; AAN40180.

XX DNA segment contg. GAL1 promoter linked to gene - useful for  
XX direction of expression of the gene in yeast cell

PS Example: Table 4, Page 21-23; 35pp; English.

CC The inventors claim a DNA segment contg. GAL1 promoter linked to  
CC gene - useful for direction of expression of the gene in yeast cell.  
CC The recombinant material carrying a GAL1 promoter of the yeast  
CC galactokinase gene may be used in expressing a desired protein, esp.  
CC bovine growth hormone, interferon, proteinin or preprorenin, in the  
CC yeast cell. Strains of *Saccharomyces cerevisiae* producing the  
CC polypeptides are produced. Yeast strains deposited as ATCC 20643,  
CC 20661, 20662 and 20663, strain designations CGY 196, 457, 461 and  
CC 528, resp. are new.

XX Sequence 381 AA;

Query Match 93.0%; Score 1919; DB 5; Length 381;

Best Local Similarity 98.4%; Pred. No. 3.5e-173; Mismatches 4; Indels 0; Gaps 0;

Matches 364; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 21 AVTHAAETRTPLRYKGSRLKALKEHGLEDFLOKQOYGISSKYSGFEVASVPLTNYD 80  
DB 12 alsqgaeltirplykgsrlkalkenhgllfedlqkqyissskysgfgevasvpltnyld 71  
QY 81 SOYFKITLGPPOEFYVLFDTGSSDFWVPSIYCKSNACKNHQRPDKRSTFQNLKPL 140  
DB 72 sqyfgkilylgrppgeftvlfdtgssdfwvpsiycksnacknhqrfdkrstfqnlgkpl 131  
QY 141 STHYTGSMOGILGYDTWTVSNIYDIOQTGSLSTQEPGDVFTYAEFGDILGMAYPSLASE 200  
DB 132 slhygtcsmgqllgydclvtvsnlvdldqvtgslstqepgdvftyaefdgllgmayspslase 191  
QY 201 YSIPVFDMMNRHLVAODLFSVYMDRNGQESMLTGAIDPSYTYGSLHWVPVTVQOYMOF 260  
DB 192 yslpfdmmnrhlvaqdlfsvymdrngesmltgaidspsytgslhwvptvqywmof 251  
QY 261 TVDSTTIGVYVACGGCOAILDGTSTKLVGPSSDILNIQOATGATQNOYGEFDIDCNL 320  
DB 252 tvdsstlgsvvaceggcailldgtstklvgpsdillnigatgatlqnydefididcdnl 311  
QY 321 SYMPVVEINGKMYPLTPSAYTSODGFCSTGFQSENHSOKWILGDVFTREYYSVFDRA 380  
DB 312 symptvveingkmypitpsaytsodgfcstgfqsenhsqkwllgdvftreyysvfdra 371  
QY 381 NNLVGLAKAI 390  
DB 372 nmlvglakai 381

RESULT 7

AAP94144  
ID AAP94144 standard; protein; 365 AA.

XX AAP94144;

XX 07-JUN-1990 (first entry)

XX Prochymosin.

XX Lactic acid bacteria: cheese; streptococcus cremoris SK112;

XX proteolase; psk112; chymosin; prochymosin.

XX NL8701378-A.

XX 02-JAN-1989.

XX 12-JUN-1987; 87NL-0001378.

XX 12-JUN-1987; 87NL-0001378.

XX (NEZU-) NEDERL INS ZUIVELON.

XX Simons AFM, De Vos WM;

DR WPI; 1989-030097/04.

DR N-PSDB; AAN91157.

PT DNA fragment having region specific for lactic acid bacteria -

PT is contained in plasmid in microorganism used in prodn. of

PT protein and food prodn. eg cheese.

PS Disclosure; 43pp; Dutch.

CC The DNA encoding prochymosin can be cloned into a plasmid ( esp. from

CC *S. cremoris* SK112) and used to produce large amts of the protein by

CC recombinant DNA techniques. This could overcome shortages of prochymosin

CC due to a shortage of calf stomachs and increasing cheese prodn.

CC Prochymosin is also used in prodn. of yoghurt, butter and buttermilk.

CC See also AAP94145 and AAP94146.

XX Sequence 365 AA;

Query Match 92.9%; Score 1918; DB 10; Length 365;

Best Local Similarity 99.7%; Pred. No. 4.1e-173; Mismatches 1; Indels 0; Gaps 0;

Matches 364; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 26 AETRIPIYKGSRLKALKEHGLEDFLOKQOYGISSKYSGFEVASVPLTNYLDSQYRG 85  
DB 1 aetirpiykgsrlkalkenhgllfedlqkqyissskysgfgevasvpltnyldsqtyfg 60  
QY 86 KIYLGTPPOEFYVLFDTGSSDFWVPSIYCKSNACKNHQRPDKRSTFQNLKPLSIHNG 145  
DB 61 kiylgtppgeftvlfdtgssdfwvpsiycksnacknhqrfdkrstfqnlgkplsihng 120  
QY 146 TSGSMOGILGYDTWTVSNIYDIOQTGSLSTQEPGDVFTYAEFGDILGMAYPSLASEYSPV 205  
DB 121 tsgsmgllgydclvtvsnlvdldqvtgslstqepgdvftyaefdgllgmayspslaseyslv 180  
QY 206 FDNMMNRHLVAODLFSVYMDRNGQESMLTGAIDPSYTYGSLHWVPVTVQOYMOFTVDSV 265  
DB 181 fdmmnrhlvaqdlfsvymdrngesmltgaidspsytgslhwvptvqywmof 240  
QY 266 TISGVVACGGCOAILDGTSTKLVGPSSDILNIQOATGATQNOYGEFDIDCNLSYMP 325  
DB 241 tisgvvaceggcailldgtstklvgpsdillnigatgatlqnydefididcdnlsympt 300  
QY 326 VVEINGKMYPLTPSAYTSODGFCSTGFQSENHSOKWILGDVFTREYYSVFDRA 385  
DB 301 vveingkmypitpsaytsodgfcstgfqsenhsqkwllgdvftreyysvfdra 360  
QY 386 LAKAI 390  
DB 361 lakai 365

RESULT 8

AAP30446  
ID AAP30446 standard; protein; 381 AA.

XX AAP30446;

XX 03-AUG-1992 (first entry)

XX Sequence encoded by preprochymosin cDNA.

XX Milk-clotting; cheese making; enzyme; zymogen.

XX Cow.

XX Location/Qualifiers

XX 1..16 /label= preprochymosin

XX 17..58 /label= prochymosin

XX 59..381 /label= chymosin

XX GB2100737-A.  
PN  
XX  
XX 06-JAN-1983.  
PD  
XX  
XX 11-JUN-1982; 82GB-0017096.  
PF  
XX  
XX 10-FEB-1982; 82GB-0003907.  
PR  
XX  
XX (CELL-) CELTECH LTD.  
PA  
XX  
PI Carey NH, Harris TUR, Lowe PA, Doel MT, Emtage JS.  
XX  
XX WPI; 1983-00545K/01.  
DR  
DR N-PSDB; AAN30209.  
XX  
XX prodn. or calf stomach chymosin for cheese making - by  
PF cultivation of micro-organisms transformed with vector system  
PS  
XX  
XX Claim 41; Fig 4; 26pp; English.  
CC  
CC The inventors claim a method for the prodn. of calf stomach chymosin  
CC for cheese making. Genes and polypeptides for preprochymosin,  
CC prochymosin and chymosin are claimed, as are vector systems and a  
CC prochymosin primer.  
XX  
XX Sequence 381 AA;

Query Match 92.9%; Score 1918; DB 4; Length 381;  
Best Local Similarity 98.4%; Pred. No. 4.3e-173;  
Matches 364; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 21 AVTHAAEITRIPLYKGSIRKALKEHGLEDFLOKQOYGISSKYSGFGEVASVPLTNYLD 80  
|::|  
DB 12 alsqgaeltiriplykgsirkaikhegllledflqkqygisksygfgevasvpltnyld 71  
OY 81 SOYFGKITLGPPEFTVLFDTGSSDFWVPSIYCKSNACKNHQRPDPKRSSTFFQNLGKPL 140  
|  
DB 72 sqyfgkiylgtrpgeftvlfdtgssdfwvpsiycksnacknhqrdprkxstffqnlgkpl 131  
OY 141 SIHYGTSGMOGLIGDVTYVSNIVDIOQTGSLSTQEPGDFVTFYAEFDGILGMAVPSLASE 200  
|  
DB 132 slhygtsgmgllgydltvtsniviqlgtvlgstqepgdfvtfyaeftdglimgavpslase 191  
OY 201 YSIPEFDNMNMNRHLVAODLFSVYMDRNGOESMLTGAIDPSYTTGSLHWPVTVQOYWQF 260  
|  
DB 192 ysipefdnmnmnrhlvaqdlfsymdringdesmltlgalnpsyytgsllhwpvrtvqywgf 251  
OY 261 TVDSVTISGVVACGCGCAIIDDGTGSKLVGPSSDILNIQAIAGATQNOYGEFDIDCNL 320  
|  
DB 252 tvdsvtisgvvacegscgaalldgtgsklvgpsdillnqaiagatqnygefdidcnl 311  
OY 321 SYMPVTVFEINCKMPLPSPATSDOGFCTSGFQSENHSCWILGDVFIREFYVSVPDRA 380  
|  
DB 312 symptvtfvtingkmyplpsaytsdqgfcstsgfqsenshsgkwllgvdvfiireyvsvfda 371  
OY 381 NNVLGLAKAI 390  
|  
DB 372 nnlvglakai 381

RESULT 9  
AAP20038  
ID AAP20038 standard; Protein: 380 AA.  
XX  
XX AAP20038;  
AC  
XX  
XX 16-DEC-1992 (first entry)  
DT  
XX  
XX Pre-prorennin-A protein sequence.  
DE  
XX  
XX Pre-pro-rennin; rennin; prorennin; enzyme; EC-3.4.23.4; chymosin;  
KW

KW protease; milk-clotting enzyme; ss.  
XX  
XX Bos taurus.  
OS  
XX  
XX GB2091271-A.  
PN  
XX  
XX 28-JUL-1982.  
PD  
XX  
XX 15-JAN-1982; 82GB-0001120.  
PF  
XX  
XX 01-DEC-1981; 81US-0325481.  
PR  
XX  
XX 16-JAN-1981; 81US-0225717.  
PR  
XX  
XX (COLB ) COLLABORATIVE RES INC.  
PA  
XX  
XX Alford BL, Mao J, Moir DT;  
PI  
XX  
XX WPI; 1982-62028E/30 (62028E).  
DR  
DR P-PSDB; AAP20038.  
XX  
XX  
XX Transformed cells producing rennin and its precursors - contg.  
PF appropriate recombinant DNA material  
PS  
XX  
XX Disclosure: Table 1; 39pp; English.  
CC  
XX  
XX DNA sequences either side of the protein sequence  
CC can be removed and are not essential to use of the gene in  
CC expression. The protein may be expressed in E. coli using  
CC plasmid pCGE21. The resulting expressed enzyme is a well  
CC known milk-clotting enzyme used in cheese-making.  
XX  
XX Sequence 380 AA;

Query Match 92.7%; Score 1914; DB 3; Length 380;  
Best Local Similarity 98.1%; Pred. No. 1e-172;  
Matches 363; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 21 AVTHAAEITRIPLYKGSIRKALKEHGLEDFLOKQOYGISSKYSGFGEVASVPLTNYLD 80  
|::|  
DB 11 alsqgaeltiriplykgsirkaikhegllledflqkqygisksygfgevasvpltnyld 70  
OY 81 SOYFGKITLGPPEFTVLFDTGSSDFWVPSIYCKSNACKNHQRPDPKRSSTFFQNLGKPL 140  
|  
DB 71 sqyfgkiylgtrpgeftvlfdtgssdfwvpsiycksnacknhqrdprkxstffqnlgkpl 130  
OY 141 SIHYGTSGMOGLIGDVTYVSNIVDIOQTGSLSTQEPGDFVTFYAEFDGILGMAVPSLASE 200  
|  
DB 131 slhygtsgmgllgydltvtsniviqlgtvlgstqepgdfvtfyaeftdglimgavpslase 190  
OY 201 YSIPEFDNMNMNRHLVAODLFSVYMDRNGOESMLTGAIDPSYTTGSLHWPVTVQOYWQF 260  
|  
DB 191 ysipefdnmnmnrhlvaqdlfsymdringdesmltlgalnpsyytgsllhwpvrtvqywgf 250  
OY 261 TVDSVTISGVVACGCGCAIIDDGTGSKLVGPSSDILNIQAIAGATQNOYGEFDIDCNL 320  
|  
DB 251 tvdsvtisgvvacegscgaalldgtgsklvgpsdillnqaiagatqnygefdidcnl 310  
OY 321 SYMPVTVFEINCKMPLPSPATSDOGFCTSGFQSENHSCWILGDVFIREFYVSVPDRA 380  
|  
DB 311 symptvtfvtingkmyplpsaytsdqgfcstsgfqsenshsgkwllgvdvfiireyvsvfda 370  
OY 381 NNVLGLAKAI 390  
|  
DB 371 nnlvglakai 380

RESULT 10  
AAR05080  
ID AAR05080 standard; protein: 365 AA.  
XX  
XX AAR05080;  
AC  
XX



DT 05-OCT-1990 (first entry)  
 XX Sequence of calf pro-rennin.  
 DE  
 XX Pro-rennin; ds.  
 KM  
 XX Bos taurus.  
 OS  
 XX JF02109984-A.  
 PN  
 XX 23-APR-1990.  
 PD  
 XX 01-JAN-1988; 88JP-0302176.  
 PF  
 XX 01-JAN-1988; 88JP-0302176.  
 PR  
 XX 01-JAN-1988; 88JP-0302176.  
 PA  
 XX (VEPP ) VEPPU T.  
 PI  
 XX WPI; 1990-168358/22.  
 DR  
 XX N-PSDB; AAQ04683.  
 PT  
 XX Complex plasmid and microbe - contains calf pro-rennin cDNA.  
 PS  
 XX Disclosure; 32; 13pp; Japanese.  
 CC  
 XX Protein product may be expressed in E.coli expression system from  
 CC plasmid pBR322.  
 CC  
 XX  
 SQ Sequence 365 AA;

Query Match 92.5%; Score 1910; DB 11; Length 365;  
 Best Local Similarity 99.2%; Pred. No. 2.3e-172;  
 Matches 362; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 26 AETRIPLYKGSLSRKALKEHGLEDFLQKQOYGISSKYSGGEVAVPLTNYLDSQYFG 85  
 DB 1 aetriplykgslsrkalkenhgllfedflqkqgylrskysgfgevavpltnyldsdyfg 60  
 QY 86 KIYLGTPQPEFTVLEFDTGSSDFWVPSIYCKSNACKNHORFDPKRSSTFONLAKPLSIHYG 145  
 DB 61 kiygtpqpeftvlfdtgssdfwpslycksnacknhorfdpkrsstfqnlgkplsihyg 120  
 QY 146 TGSMOGILGYDTVTYNSNIVDIQOTVGLSTOEPGDFVTFYAEFDGIAGMAYPSLASYSIPV 205  
 DB 121 tgsmgilgydtvtvnsnivdiqotvglstqepgdfvtfyaefdgllgmaypslaseysipv 180  
 QY 206 FDNMNNRHLVAODLFESVYMDRNGESMLTLGAIDPSYTTGSLHWVPVTVOQYWFQTVDSV 265  
 DB 181 fdnmnnrhlvaodlfesvymdrngesmltlgaidsytslhwvpvtvqywfqtvdsav 240  
 QY 266 TISGVVAVCEGCGAIIIDTGTSKLVGPSDDLINIOAIGARONOGEDDIDCDNISYPT 325  
 DB 241 tisgvvavceggcgaiiildgtstklvgpsddlinaigaarqngydeididcdnisympt 300  
 QY 326 VFEINGKMYPLTPSAVTSODGECTSGFQSEHNSOKWILGDVFTREYYSVFDRANNLV 385  
 DB 301 vfeingkmyp ltpsaytsodgectsgfsgenhsqkwilgdvftreyysvfdrannlvg 360  
 QY 386 LAKAI 390  
 DB 361 laktl 365

RESULT 11  
 AAP30603  
 ID AAP30603 standard; protein; 365 AA.  
 XX  
 AC AAP30603;  
 XX  
 DT 14-JUN-1992 (first entry)  
 AC  
 XX Sequence encoded by prorennin cDNA in PCR 10001.

XX Rennin; renin; enzyme; protease.  
 KM  
 XX Bos taurus.  
 OS  
 XX EP73029-A.  
 PN  
 XX 02-MAR-1983.  
 PD  
 XX 19-AUG-1982; 82EP-0107601.  
 PF  
 XX 24-AUG-1981; 81JP-0131631.  
 PR  
 XX (BEPP/) BEPPU T.  
 PA  
 XX Beppu T, Uozumi T, Nishimori K;  
 PI  
 XX WPI; 1983-22976K/10.  
 DR  
 XX N-PSDB; AAN30063.  
 PT  
 XX Plasmid contg. calf pro:rennin DNA - and transformed  
 PT microorganisms  
 PS  
 XX Example; Page 20-23; 32pp; English.  
 CC  
 XX The inventors claim recombinant plasmids contg. the cDNA of calf  
 CC prorennin. Specified plasmids are PCR 10001 and PCR2001 (contg. the  
 CC whole sequence plus the lac promoter region). Also new are  
 CC microorganisms transformed with the plasmids, esp. E. coli CRI (ATCC  
 CC 391710) contg. plasmid PCR2001.  
 CC  
 XX  
 SQ Sequence 365 AA;

Query Match 92.4%; Score 1907; DB 4; Length 365;  
 Best Local Similarity 98.9%; Pred. No. 4.5e-172;  
 Matches 361; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 26 AETRIPLYKGSLSRKALKEHGLEDFLQKQOYGISSKYSGGEVAVPLTNYLDSQYFG 85  
 DB 1 aetriplykgslsrkalkenhgllfedflqkqgylrskysgfgevavpltnyldsdyfg 60  
 QY 86 KIYLGTPQPEFTVLEFDTGSSDFWVPSIYCKSNACKNHORFDPKRSSTFONLAKPLSIHYG 145  
 DB 61 kiygtpqpeftvlfdtgssdfwpslycksnacknhorfdpkrsstfqnlgkplsihyg 120  
 QY 146 TGSMOGILGYDTVTYNSNIVDIQOTVGLSTOEPGDFVTFYAEFDGIAGMAYPSLASYSIPV 205  
 DB 121 tgsmgilgydtvtvnsnivdiqotvglstqepgdfvtfyaefdgllgmaypslaseysipv 180  
 QY 206 FDNMNNRHLVAODLFESVYMDRNGESMLTLGAIDPSYTTGSLHWVPVTVOQYWFQTVDSV 265  
 DB 181 fdnmnnrhlvaodlfesvymdrngesmltlgaidsytslhwvpvtvqywfqtvdsav 240  
 QY 266 TISGVVAVCEGCGAIIIDTGTSKLVGPSDDLINIOAIGARONOGEDDIDCDNISYPT 325  
 DB 241 tisgvvavceggcgaiiildgtstklvgpsddlinaigaarqngydeididcdnisympt 300  
 QY 326 VFEINGKMYPLTPSAVTSODGECTSGFQSEHNSOKWILGDVFTREYYSVFDRANNLV 385  
 DB 301 vfeingkmyp ltpsaytsodgectsgfsgenhsqkwilgdvftreyysvfdrannlvg 360  
 QY 386 LAKAI 390  
 DB 361 laktl 365

RESULT 12  
 AAP30013  
 ID AAP30013 standard; peptide; 379 AA.  
 XX  
 AC AAP30013;  
 XX



Db 312 symplvvfeingkmypitpsaytsqdgqfctsaftsenhsqkwlvgdvfireyvsfdra 371  
 QY 381 NNVLGLAKAI 390  
 |||||||||  
 Db 372 nnlvglakal 381

## RESULT 14

AAR20730  
 ID AAR20730 standard; protein: 381 AA.  
 AC AAR20730;  
 XX 13-MAY-1992 (first entry)  
 DT  
 DE Prochymosin (prorennin).  
 XX  
 KM Rennet; zymogen; autocatalytic activation; pseudorennin; milk;  
 KM clotting activity; renin A, rennin B.  
 XX  
 OS Bos taurus.  
 PH Key Location/Qualifiers  
 FT Protein 1..381  
 FT /label= preprorennin  
 FT /note= "also known as preprochymosin"  
 FT Peptide 1..17  
 FT /label= signal\_peptide  
 FT Cleavage-site 43..44  
 FT /note= "autocatalytic cleavage at pH 2"  
 FT Cleavage-site 59..60  
 FT /note= "autocatalytic cleavage pH 4.7"  
 FT MISC-difference 306  
 FT /note= "amino acid determining A or B form of rennin"  
 FT MISC-difference 218  
 FT /note= "Asp in published sequence"  
 FT  
 PN US5082775-A.  
 XX 21-JAN-1992.  
 PD  
 XX 28-OCT-1988; 88US-0263927.  
 PF  
 XX 28-OCT-1988; 88US-0263927.  
 PR 11-MAY-1984; 88US-0263927.  
 PR 12-DEC-1986; 88US-0940195.  
 PR 31-MAR-1983; 83US-0480860.  
 PR 28-APR-1986; 86US-0856700.  
 XX  
 PA (BERL-) BERLEX LABS INC.  
 XX  
 PI McCaman MT, King JF;  
 DR MPI: 1992-049149/06.  
 DR N-PSDB: AAQ20949.  
 XX  
 PT Isolating heterologous polypeptide from bacterial inclusion  
 PT bodies - by lysing cells, extrn. with nonionic detergent and sepg.  
 PT Insoluble polypeptide  
 XX  
 PS Disclosure; Fig 6; 21pp; English.  
 XX  
 CC The prorennin (prochymosin) sequence was deduced from the DNA  
 CC sequence obtd. by screening a cDNA library with rennin specific  
 CC probes derived from the published amino acid sequence of rennin  
 CC (Poltmann et al., J. Biol. Chem. 254, 8447-8456 (1979)) (see 223291.2).  
 CC The deduced protein sequence was the same as the published sequence  
 CC after cleavage of the signal peptide, except for an amino acid change  
 CC at residue 218, which may be an artefact of protein sequencing in  
 CC the original sequence. The cleavage product of prorennin, rennin is  
 CC synthesized in two active forms rennin A, and rennin B. The cleaved  
 CC secreted protein undergoes activation in the acidic conditions of the

CC stomach to generate mature rennin by two autocatalytic cleavage events.  
 CC Rennin is an active component of rennet which is used to clot milk in  
 CC the process of making cheese.  
 CC See also AAR22417.  
 XX  
 SO Sequence 381 AA;

Query Match 91.5%; Score 1888; DB 13; Length 381;  
 Best Local Similarity 97.3%; Pred. No. 3e-170;  
 Matches 361; Conservative 4; Mismatches 4; Indels 2; Gaps 2;

QY <1 AVTHAAETRIPILYGKSLRKALKEHGLEDFLQKQYGISKYSGEFVASVPLTNLD 80  
 :: |||||||  
 Db 12 alsqgaeltirpilykgsrlrkalkenhglldedflqkqgisksygsfgevsvpltnlyld 71  
 QY 81 SQYFGKIYLGTPPOEFVTLFDTGSSDFWVPSIYCKSNACKNHQRPDKSSTFONLCKPL 140  
 Db 72 sqyfgkiylgtppegeftvlfdtgssdfwvpslycksnacknhqrfdkpsstfqnlgkpl 131  
 QY 141 SIHYGTGSMOGLIGVDPTVTSNIVDIOQTGVLSTQEPDVTTFYAEFDGILGMAY-PSIAS 199  
 Db 132 sihygtgsmgylgldvtvsnivdigeftvglstqepdvtflyaeftdglgmaypslas 191  
 QY 200 EYSIPVEFDNMNRRHLVADLFVSVMWRNGQESMLTGAIDPSYTGSLHWVPVYQOXYWQ 259  
 Db 192 eysipvfnnm-rhlvagdlsfymdrngqesmltgaidsytsfslhwvprvtyqywg 250  
 QY 260 FTVDVTSISGVVACEGGQAILDGTGSKLVGPSSDIINIOAIGATONOGYGEFDIDCDN 319  
 Db 251 ftvdvtsisgvvacdgccqaildgtgsklvgpsdlinlqgaltqngydefdidcdn 310  
 QY 320 LSYMPTVYFEINGKMYRLTPSAYTSQDGRCTSGQSQSNHSOKMTLGDVFTREYYSVDR 379  
 Db 311 lsymptvfeingkmypitpsaytsqdgqfctsaftsenhsqkwlvgdvfireyvsfdr 370  
 QY 380 ANNVLGLAKAI 390  
 |||||||||  
 Db 371 annlvglakal 381

RESULT 15  
 AAP94370  
 ID AAP94370 standard; protein: 450 AA.  
 XX  
 AC AAP94370;  
 XX  
 DT 15-JUN-1990 (first entry)  
 DT  
 XX  
 DE Sequence encoded by BamHI insert from pDM100PC.  
 XX  
 KM Kluyveromyces; pDM100PC; chymosin; tissue plasminogen activator.  
 XX  
 PN EP301670-A.  
 PD  
 XX 01-JAN-1989.  
 PF  
 XX 28-JUL-1988; 88EP-0201632.  
 PR 28-JUL-1987; 87US-0078539.  
 XX  
 PA (KONN ) GIST-BROCADES NV.  
 XX  
 PI van den Berg JA, van Ooyen AJJ, Rietveld K;  
 DR MPI: 1989-033565/05.  
 DR N-PSDB: AAN91185.  
 XX  
 PT Kluyveromyces host cells for producing polypeptide(s) -  
 PT used for highly efficient prodn. of eg chymosin tissue  
 PT plasminogen activator or human serum albumin.  
 XX  
 PS Disclosure; : 56pp; English.

XX The PDM100PC sequence product comprises a fusion peptide of the alpha-  
 CC factor of *S.cerevisiae* and prochimysin and transcriptional and regulatory  
 CC control regions. The Kluyveromyces expression systems provide highly  
 CC efficient secretion and processing of a wide variety of proteins.  
 CC Sequences identical to those published in EP301669.

XX  
 SQ Sequence 450 AA:

Query Match 90.8%; Score 1874; DB 10; Length 450;  
 Best Local Similarity 95.7%; Pred. No. 8.2e-169;  
 Matches 355; Conservative 5; Mismatches 11; Indels 0; Gaps 0;

QY	20	VAVTHAAETTRIPLYKGSLSRKALKENGLLEDPLQKQYGISSKYSGFGEVASYPLTNYL	79
		:	
Db	80	vsldkreaitrplykgskslkalkehgllledflqkqgysksygsfgevavpltnyp	139
QY	80	DSQYFGKIYLGTPPOEFTVLEFDTGSSDFWVPSIYCKSNACKNHQRPDRKSTFQNLGKP	139
Db	140	dsqyfgkiyigtppqetvlfdtgssdfwvpsmycksnacknhgrfdprkstftqnlgkp	199
QY	140	LSIHGSGSMGILIGYDVTYNSNIVDIQQTWGLSTOEPGDVETVAEPDGLGMAYPSLAS	199
Db	200	lsihgsgsmgillgydvtvsnlvdliqvtvlgstdepdvftyaetdgllgmaypslas	259
QY	200	EYSIPVFDNMNRRHLVAQDLFSVYMDRNGQESMLTLGADPSYXTGSLHWVPVTVOQYMQ	259
Db	260	eystsvfdnmnnrhlvagdflsvymdrngqesmltlgpidpsyxtgslhwvptvqywg	319
QY	260	FTVDSVITISGVVACBEGCCAILDTGTSKLVGPSDDIINTIOAIGATONOGEDIDCDN	319
Db	320	ftvdsvtisgvvpoeggcgailldtgtsklvgpsddlnlqplgatnqygdfddicdn	379
QY	320	LSYPTVVFEEINGKMWPLTPSAYTSODGFCFSGFOSENHSSOKNLTGDFIREYYSVFDK	379
Db	380	lsyptvtfvfeingkiypltpsaaytsodgfcfsgfgeenhsgkwlldgdfireyysvfdk	439
QY	380	ANNLVGLAKAT	390
Db	440	pnnlvglakal	450

Search completed: July 31, 2002, 09:20:52  
 Job time: 1554 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 31, 2002, 09:16:08 ; Search time 20.9 seconds  
(without alignments)  
1793.055 Million cell updates/sec

Title: US-09-643-755B-2  
Perfect score: 2064  
Sequence: 1 MNFLKSPFVAFICPGQYFV.....REYVSVPDRANLVGLAKAI 390

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 segs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1932	93.6	381	1	chymosin (EC 3.4.2
2	1829	88.6	381	1	chymosin (EC 3.4.2
3	1605	77.8	381	2	chymosin (EC 3.4.2
4	1601	77.6	380	2	chymosin (EC 3.4.2
5	1108.5	53.7	387	2	chymosin (EC 3.4.2
6	1099.5	53.3	385	2	chymosin (EC 3.4.2
7	1094.5	53.0	383	2	chymosin (EC 3.4.2
8	1092.5	52.9	388	1	chymosin (EC 3.4.2
9	1085.5	52.6	388	1	chymosin (EC 3.4.2
10	1084.5	52.5	388	1	chymosin (EC 3.4.2
11	1082.5	52.4	388	1	chymosin (EC 3.4.2
12	1080.5	52.3	388	1	chymosin (EC 3.4.2
13	1078.5	52.3	388	1	chymosin (EC 3.4.2
14	1075.5	52.1	388	1	chymosin (EC 3.4.2
15	1074	52.0	384	2	chymosin (EC 3.4.2
16	1070	51.8	386	1	chymosin (EC 3.4.2
17	1045	50.6	387	2	chymosin (EC 3.4.2
18	1037	50.2	387	2	chymosin (EC 3.4.2
19	1034	50.1	387	2	chymosin (EC 3.4.2
20	1030	49.9	387	2	chymosin (EC 3.4.2
21	1016.5	49.2	382	1	chymosin (EC 3.4.2
22	957.5	46.4	334	2	chymosin (EC 3.4.2
23	942.5	45.1	339	2	chymosin (EC 3.4.2
24	930	45.1	339	2	chymosin (EC 3.4.2
25	923	44.7	336	2	chymosin (EC 3.4.2
26	918	44.5	336	2	chymosin (EC 3.4.2
27	911.5	44.2	331	2	chymosin (EC 3.4.2
28	894	43.3	333	2	chymosin (EC 3.4.2
29	892	43.2	332	1	chymosin (EC 3.4.2

30	891.5	43.2	384	2	chymosin (EC 3.4.2
31	876	42.4	377	1	chymosin (EC 3.4.2
32	869	42.1	388	2	chymosin (EC 3.4.2
33	866	42.0	394	2	chymosin (EC 3.4.2
34	862	41.8	388	2	chymosin (EC 3.4.2
35	859	41.6	376	2	chymosin (EC 3.4.2
36	843.5	40.9	389	2	chymosin (EC 3.4.2
37	826.5	40.0	365	2	chymosin (EC 3.4.2
38	818.5	39.7	387	2	chymosin (EC 3.4.2
39	797.5	38.6	398	2	chymosin (EC 3.4.2
40	792.5	38.4	412	1	chymosin (EC 3.4.2
41	784.5	38.0	410	1	chymosin (EC 3.4.2
42	775	37.5	407	1	chymosin (EC 3.4.2
43	747.5	36.2	389	2	chymosin (EC 3.4.2
44	744	36.0	380	2	chymosin (EC 3.4.2
45	728.5	35.3	344	1	chymosin (EC 3.4.2

## ALIGNMENTS

RESULT 1

CMBO

chymosin (EC 3.4.23.4) precursor - bovine

N:Alternate names: prechymosin; preprorennin; rennin B

N:Contents: chymosin; prochymosin

C:Species: Bos primigenius taurus (cattle)

C>Date: 24-Apr-1984 #sequence-revision 09-Sep-1994 #text-change 18-Jun-1999

C:Accession: A25631; A93419; A44608; A92259; A91935; A91495; D22434; A00985

R:Hisada, M.; Sasaki, K.; Uozumi, T.; Beppu, T.

Gene 43, 197-203, 1986

A:Title: Cloning and structural analysis of the calf prochymosin gene.

A:Reference number: A25631; MUID:86301873

A:Accession: A25631

A:Molecule type: DNA

R:Hisada, M.; Sasaki, K.; Uozumi, T.; Beppu, T.

Nucleic Acids Res. 10, 2177-2187, 1982

A:Title: Molecular cloning and nucleotide sequence of cDNA coding for calf preprochym

A:Reference number: A93419; MUID:82221400

A:Contents: prochymosin B

A:Accession: A93419

A:Molecule type: mRNA

A:Residues: 1-229, 'N', 231-381 <HAR>

A:Cross-references: GB:J00003; NID:g162859; PID:AAA30448.1; PID:g162860

R:Nishimori, K.; Kawaguchi, Y.; Hisada, M.; Uozumi, T.; Beppu, T.

J. Biochem. 91, 1085-1088, 1982

A:Title: Nucleotide sequence of calf prorennin cDNA cloned in Escherichia coli.

A:Reference number: A44608; MUID:82189915

A:Accession: A44608

A:Molecule type: mRNA

A:Residues: 7-14, 'X', 16-42, 'L', 44-87, 'N', 89-301, 'D', 303-324, 'I', 326-334, 'G', 336-342, 'A:Cross-references: GB:J00004

A:Note: authors translated the codon TTG for residue 43 as Phe, ACC for residue 82 as S Ala

R:Polmann, B.; Pedersen, V.B.; Kauffman, D.; Wybrandt, G.

J. Biol. Chem. 254, 8447-8456, 1979

A:Title: The primary structure of calf chymosin.

A:Reference number: A92259; MUID:79239460

A:Contents: chymosin B; disulfide bonds

A:Accession: A92259

A:Molecule type: protein

A:Residues: 59-217, 'D', 219-381 <FOL>

R:Polmann, B.; Pedersen, V.B.; Jacobsen, H.; Kauffman, D.; Wybrandt, G.

Proc. Natl. Acad. Sci. U.S.A. 74, 2321-2324, 1977

A:Title: The complete amino acid sequence of prochymosin.

A:Reference number: A44620; MUID:77234648

A:Accession: A44620

A:Molecule type: protein

A:Residues: 17-173, 'T', 175-217, 'D', 219-251, 'Y', 253-381 <FO2>

R:Chang, W.J.; Takahashi, K.

J. Biochem. 76, 467-474, 1974

A:Title: The structure and function of acid proteases. III. Isolation and characteriz

A:Reference number: A91935; MWID:75060332  
A:Contents: active site peptides  
A:Accession: A91935  
A:Molecule type: protein  
A:Residues: 92-96;274-278, 280 <CHA>  
A:Note: The authors did not find Lys-279 in their active site peptide  
R:Moit, D.; Mao, J.; Schumm, J.W.; Vovis, G.F.; Alford, B.L.; Taunton-Rigby, A.  
Gene 19, 127-138, 1982  
A:Title: Molecular cloning and characterization of double-stranded cDNA coding for bovin  
A:Reference number: A91495; MWID:83054629  
A:Contents: prochymosin A  
A:Accession: A91495  
A:Molecule type: mRNA  
A:Residues: 1-301,'D','303-381 <MOI>  
R:Richihara, Y.; Sogawa, K.; Takahashi, K.  
J. Biochem. 98, 483-492, 1985  
A:Title: Isolation of human, swine, and rat prepepsinogens and calf preprochymosin, and  
A:Reference number: A22434; MWID:86059312  
A:Accession: D22434  
A:Molecule type: protein  
A:Residues: 1-18, 'XX','21, 'XX','24, 'X','26, 'X','28, 'X','30 <ICH>  
C:Comment: The sequence of variant B is shown.  
C:Comment: Chymosin is synthesized in the mucosa of the abomasum (fourth stomach) of you  
C:Comment: Forms A and B are probably allelic variants. Chymosin B is the predominant fo  
C:Genetics:  
A:Introns: 21/2; 71/3; 111/1; 150/3; 217/2; 255/2; 303/3; 336/3  
C:Superfamily: peptsin  
C:Keywords: aspartic proteinase; gastric juice; hydrolase; protein digestion; stomach; z  
F:1-16/Domain: signal sequence #status experimental <SIG>  
F:15-381/Product: activation peptide #status experimental <APT>  
F:92-274/Active site: Asp #status experimental  
F:103-110, 265-269, 308-341/Disulfide bonds: #status experimental

Query Match 93.6%; Score 1932; DB 1; Length 381;  
Best Local Similarity 98.9%; Pred. No. 6,3e-145;  
Matches 366; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

21 AYTHAAETRIPIYKGSRLKRAKKEHGLEDFLOKQOYGISSKYSQGEVASYPLTNYLD 80  
::: |||||  
12 ALSQGAETRIPIYKGSRLKRAKKEHGLEDFLOKQOYGISSKYSQGEVASYPLTNYLD 71  
81 SQYFGKIYLGTPQPEFVNLDTGSSDFWVPSTYCKSNACKNNQRPDRKRSSTQNCKPL 140  
72 SQYFGKIYLGTPQPEFVNLDTGSSDFWVPSTYCKSNACKNNQRPDRKRSSTQNCKPL 131  
141 SHYKGSQMOGILGYDPTVATSNIVDIOQTGISTQEPGVFYAEFDGLIGMAYPSLASE 200  
132 SHYKGSQMOGILGYDPTVATSNIVDIOQTGISTQEPGVFYAEFDGLIGMAYPSLASE 191  
201 YSIPVEDNNMNRHLVAODLFSVYMDRNGQESMLTLGAIDPSYYTGSLSHWVPVTVOQYWF 260  
192 YSIPVEDNNMNRHLVAODLFSVYMDRNGQESMLTLGAIDPSYYTGSLSHWVPVTVOQYWF 251  
261 YDVSVITISGVVACGEGCOALIDTGTSKLVGPSSDIILNQALIGATQNOYGEDEDICDNL 320  
252 YDVSVITISGVVACGEGCOALIDTGTSKLVGPSSDIILNQALIGATQNOYGEDEDICDNL 311  
321 SYMPTVVEELNGKMYPLTPPSATYSODQGFCTSGFQSENHSGKMIIGDVFIREYVAFDRA 380  
312 SYMPTVVEELNGKMYPLTPPSATYSODQGFCTSGFQSENHSGKMIIGDVFIREYVAFDRA 371  
381 NNLVGLAKAI 390  
372 NNLVGLAKAI 381

RESULT 2  
CMSHB  
chymosin (EC 3.4.23.4) B precursor - sheep  
C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)  
C:Date: 30-Sep-1991 #sequence, revision 30-Sep-1991 #text\_change 18-Jun-1999

C:Accession: S10996  
R:Pungercar, J.; Strükelj, B.; Gubensek, F.; Turk, V.; Kregar, I.  
Nucleic Acids Res. 18, 4602, 1990  
A>Title: Complete primary structure of lamb preprochymosin deduced from cDNA.  
A:Reference number: S10996; MUID:90356410  
A:Accession: S10996  
A:Molecule type: mRNA  
A:Residues: 1-381 <PUN>  
A:Cross-references: EMBL:X53037; NID:g1373; PIDN:CAA37209.1; PID:g1374  
C:Superfamily: pepsin  
C:Keywords: aspartic proteolase; gastric juice; hydrolase; protein digestion; stomach  
F:1-16/Domain: signal sequence #status predicted <Sig>  
F:17-58/Domain: activation peptide #status predicted <APV>  
F:59-381/Product: chymosin B #status predicted <MAT>  
F:92,274/Active site: Asp #status predicted  
F:105-110,265-269,308-341/Disulfide bonds: #status predicted

Query Match            88.6%   Score 1829; DB 1; Length 381;  
Best local similarity   93.2%; Pred. No. 8,7e-137;  
Matches   345; Conservative   13; Mismatches   12; Indels   0; Gaps   0;

DB        21 AVTHAAETTRIPLYKGSLSRKALKEHGLEDFLOKQOYGISSKYSGEFEVASVPLNYLD 80  
          ::: |::::::::::|::::::::::|::::::::::|::::::::::|::::::::::|  
12 ALSQAETRIPIPLYKGKFLRKALKERGLLEDFLOKOYGVSSESGEVAASVPLTNVLD 71

QY        81 SOYFKKIYLGPPEGFYTLFDTGSSDFWVPSTYCKSNACKNHQRPDPKSTPONLKPL 140  
          ::: |::::::::::|::::::::::|::::::::::|::::::::::|::::::::::|  
DB        72 SOYFKKIYLGPPEGFYTLFDTGSSDFWVPSTYCKSNACKNHQRPDPKSTPONLKPL 131

QY        141 SIHYGTGSMOGILGDVTYASNIVDIQOTVGLSTPOEPDVTYAEFGIILMAVPSLASE 200  
          || |::::::::::|::::::::::|::::::::::|::::::::::|::::::::::|  
DB        132 STRYGTGSMOGILGDVTYASNIVDIQOTVGLSTPOEPDVTYAEFGIILMAVPSLASE 191

QY        201 YSIPVEDNMNRHLVAODLFESYMDRNQGESMLTGLAIDPSYTYGSLHWBPVTVOQTWOQF 260  
          || |::::::::::|::::::::::|::::::::::|::::::::::|::::::::::|  
DB        192 YSVPEFDNMNRHLVAODLFESYMDRNSGGSGMLTGLAIDPSYTYGSLHWBPVTVOQTWOQF 251

QY        261 TVDSVTISGVAVACGGCGQAILDTGSTKLVPSPSSILNIQDAITATONQVEFPIDCNL 320  
          || |::::::::::|::::::::::|::::::::::|::::::::::|::::::::::|  
DB        252 TVDSVTISGVAVACGGCGQAILDTGSTKLVPSPSSILNIQDAITATONQVEFPIDCDNL 311

QY        321 SYMPVVEEINGKNMPLPPSAVTSQDGFCISGFQSENHSKWIIIGDVIFIREYVSVPDRA 380  
          || |::::::::::|::::::::::|::::::::::|::::::::::|::::::::::|  
DB        312 SSMPVVEEINGKNMPLPPSAVTSQDEGFCISGFQGENHSHQWIIIGDVIFIREYVSVPDRA 371

QY        381 NNLVGLAKAI 390  
          |||||  
DB        372>NNLVGLAKAI 381

RESULT     3  
JC7247  
prochymosin - common marmoset  
C.Species: Callithrix jacchus (common marmoset)  
C.Date: 09-Jun-2000 #sequence\_revision 09-Jun-2000 #text\_change 21-Jul-2000  
C.Accession: JC7247  
R:Kageyama, T.  
J. Biochem. 127, 761-770, 2000  
A.Title: New world monkey pepsinogens A and C, and prochymosins. Purification, charac  
A:Reference number: JC7245  
A:Accession: JC7247  
A:Molecule type: mRNA  
A:Residues: 1-381 <KAG>  
A:Cross-references: DDBJ:AB038386  
A:Experimental source: strain NM813  
C.Comment: This protein, a zymogen of pepsins, is the major proteolytic enzyme in ver  
volution of mammalian orders and families.  
C:Superfamily: pepsin  
C:Keywords: gastric juice; zymogen

Query Match            77.8%   Score 1605; DB 2; Length 381.

Best Local Similarity 81.4%, Pred. No. 4.3e-119;  
Matches 301; Conservative 30; Mismatches 39; Indels 0; Gaps 0;

```

Oy 21 ATHTAAETRIPLRYKSKLRKALKKEHGLEDFLOKQOYGISSKYSGFGEVASVPLNTYD 80
    :::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 12 ALSQASGIVRIPLHKGSKLRKALKKEHGLEDFLOKHQHAVSRKHSRSRVASFEFLTYND 71
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Oy 81 SOYFGKIYLGTPPOEFTVLEDTGSSDFWVPVSIYCKSNACKNHRPDRKSTFQNLKRP 140
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 72 COYFGKIYIGTPPOEFTVLEDTGSSDLWVPVSYCNVACNHNHRRPDRKSTFQNMKDL 131
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Oy 141 SIHYGTGSKQGLIGDTVTVSNIVDIQOTVGLSTQEPGDVFTYAEFDGILGMAVPSLASE 200
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 132 SIQYGTGSKQGLIGDTVTVSSIVDPHQOTVGLSTQEPGDVFTYAEFDGILGLAPSLASE 191
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Oy 201 YSIPEFDNMNRHLVAODLFESVYMDNRNGESMLTGAIDPSYTGSLHWPVTVQOYKMF 260
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 192 YSVPEFDNMNRHLVAODLFESVYMSRNEGSMILTGAIDPSYTGSLHWPVTVQOYKMF 251
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Oy 261 TVDSVTISGVVACGEGCOAILDGTGSKLVGPSDDLNTQAIQAIQANQYGEFDIDCDNL 320
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 252 TVDSVTYVGAVVACDGGCOAILDGTGSMLVGPSDFINQAIQAIQANQYGEFDIDCDNL 311
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Oy 321 SYMPVTVVEINCKMTPLPISAVTSODQFCYSGFQSENHSQKWIIGDVPIREYYSVFDRA 380
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 312 SSMPVTVVEINCKMTPLPISAVTNODQFCYSGFQSDSSQOUILGDVPIREYYSVFDRA 371
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Oy 381 NNVLGLAKAI 390
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 372 SNVLGLAKAI 381
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

## RESULT 4

147176  
chymosin (EC 3.4.23.4) precursor - pig (fragment)  
C:Species: Sus scrofa domestica (domestic pig)  
C:Date: 21-Feb-1997 #sequence\_revision 21-Feb-1997 #text\_change 20-Jun-2000  
C:Accession: I47176  
R:Foltmann, B.; Jensen, A.L.; Loenblad, P.; Smidt, E.; Axelssen, N.H.  
Comp. Biochem. Physiol. A 68, 9-13, 1981  
A:Title: A Developmental Analysis of the Production of Chymosin and Pepsin in Pigs.  
A:Reference number: I47176  
A:Accession: I47176  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-380 <FOL>  
A:Cross-references: EMBL:U14406; NID:g540096; PIDN:AAB08492.1; PID:g540097  
C:Superfamily: pepsin  
C:Keywords: aspartic proteinase; hydrolase

Query Match 77.6%; Score 1601; DB 2; Length 380;  
Best Local Similarity 80.1%; Pred. No. 8.8e-119;  
Matches 297; Conservative 35; Mismatches 39; Indels 0; Gaps 0;

```

Oy 20 VAVTHAAETRIPLRYKSKLRKALKKEHGLEDFLOKQOYGISSKYSGFGEVASVPLNTYL 79
    :::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 10 LALSGSGITRPLRKSKLRKALKKEHGLEDFLOKOPALSSKYSSEFEVASSEPLNTYL 69
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Oy 80 DSQYFGKIYLGTPPOEFTVLEDTGSSDFWVPVSIYCKSNACKNHRPDRKSTFQNLKRP 139
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 70 DTQYFGKIYIGTPPOEFTVLEDTGSSSELMVPSYCKSDACQNHHRFPDRKSTFQNLKRP 129
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Oy 140 LSIHYGTGSKQGLIGDTVTVSNIVDIQOTVGLSTQEPGDVFTYAEFDGILGMAVPSLAS 199
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 130 LSIQYGTGSKQGLIGDTVTVMAGIVDAHQTVGLSTQEPGDVFTYAEFDGILGLYPELAS 189
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Oy 200 EYSIPEFDNMNRHLVAODLFESVYMDNRNGESMLTGAIDPSYTGSLHWPVTVQOYKMF 259
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 190 EYTVPEFDNMNRHLVAODLFAYVMSRNDGSMILTGAIDPSYTGSLHWPVTVQOYKMF 249
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Oy 260 FTVDVSVTISGVVACGEGCOAILDGTGSKLVGPSDDLNTQAIQAIQANQYGEFDIDCDN 319
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

```

Db 250 FTVDVSVTVINGVAVACGCGOAILDGTGSMLVGPSDDLNTQAIQAIQANQYGEFDIDCGS 309
Oy 320 LSYMPVTVVEINCKMTPLPISAVTSODQFCYSGFQSENHSQKWIIGDVPIREYYSVFDRA 379
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 310 SSMPVTVVEINCKMTPLPISAVTNODQFCYSGFQSDSSQOUILGDVPIREYYSVFDRA 369
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Oy 380 NNVLGLAKAI 390
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 370 SNVLGLAKAI 380
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

## RESULT 5

JC7245  
pepsinogen A - common marmoset  
C:Species: Callithrix jacchus (common marmoset)  
C:Date: 09-Jun-2000 #sequence\_revision 09-Jun-2000 #text\_change 21-Jul-2000  
C:Accession: JC7245  
R: Kageyama, T.  
J. Biochem. 127, 761-770, 2000  
A:Title: New world monkey pepsinogens A and C<sub>1</sub> and prochymosins. Purification, charac  
A:Reference number: JC7245  
A:Accession: JC7245  
A:Molecule type: mRNA  
A:Residues: 1-387 <KAG>  
A:Cross-references: DBJ:AB038384  
A:Experimental source: strain NM794  
C:Comment: This protein, a zymogen of pepsins, is the major proteolytic enzyme in ver  
volution of mammalian orders and families.  
C:Superfamily: pepsin  
C:Keywords: gastric juice; zymogen

Query Match 53.7%; Score 1108.5; DB 2; Length 387;  
Best Local Similarity 58.6%; Pred. No. 7.3e-80;  
Matches 215; Conservative 49; Mismatches 94; Indels 9; Gaps 3;

```

Oy 30 RIPLYKSKLRKALKKEHGLEDFLOKQOYGISSKYSGFGE---VASVPLNTYLDISQYFG 85
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 18 KVSILKKSKLRKALKKEHGLEDFLOKKNITIDPASKYFPOGEAATMINOPLVNTLMEYFG 77
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Oy 86 KIYLGTPPOEFTVLEDTGSSDFWVPVSIYCKSNACKNHRPDRKSTFQNLKPLSIHYG 145
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 78 TIGIGTPPOEFTVLEDTGSSNLMVPSIYCSPPACTNHNFRNPQESSTYATSTLSIAYG 137
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Oy 146 TSGMSGILGDTVTVSNIVDIQOTVGLSTQEPGDVFTYAEFDGILGMAVPSLASESTIV 205
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 138 TGSMTGILGDTVTVOYGGIDVNOIFELSETERPSFLYSPFBDIGLAVPSISSSGATPV 197
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Oy 206 FDNMNRHLVAODLFESVYMDNRNGOE-SMLTGAIDPSYTGSLHWPVTVQOYKMFYVDS 264
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 198 FDNIMNODLVSDPLSVYSSNDQSGSVYMFPGIDBSYTTGSLHWPVVAEGTWQITVDS 257
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Oy 265 VYISGVVACGEGCOAILDGTGSKLVGPSDDLNTQAIQAIQANQYGEFDIDCDNLISYMP 324
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 258 IIMNGEALICACGCOAIVDTGTSILSGPSPAIANTIOSYIGASENSGEMVWVSCSAISSLP 317
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Oy 325 TVFEELNGKMYPLTPISAVTSODQFCYSGFQSEN----HSQKWIIGDVPIREYYSVFDRA 380
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 318 DIVFTINGIYVPAASATYLLDDEGGCTSGFQGNIPFAYGELMILGDVPIROYFAVFDRA 377
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Oy 381 NNVLGLA 387
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 378 NNQVGLA 384
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

## RESULT 6

JC7575  
pepsinogen A - bullfrog  
C:Species: Rana catesbeiana (bullfrog)  
C:Date: 30-Jun-2001 #sequence\_revision 30-Jun-2001 #text\_change 03-Aug-2001  
C:Accession: JC7575  
R: Ikuzawa, M.; Inokuchi, T.; Kobayashi, K.; Yasumasu, S.  
J. Biochem. 129, 147-153, 2001







```

Db          377  RANNOVGILA 385
|||||
|||||
RESULT      11
PEHU
pepsin A (EC 3.4.23.1) 3 precursor [validated] - human
N:Alternate names: pepsinogen A Isozyme 3
C:Species: Homo sapiens (man)
C:Date: 19-Feb-1984 #sequence_revision 19-Feb-1984 #text_change 08-Dec-2000
C:Accession: A00960; PX0023; S02663; F22434; I54252; PX0024
R:Sogawa, K.; Fujii-Kuriyama, Y.; Mizukami, Y.; Ichihara, Y.; Takahashi, K.
J. Biol. Chem. 258, 5306-5311, 1983
A>Title: Primary structure of human pepsinogen gene.
A:Reference number: A00960; MUID:83161158
A:Accession: A00960
A:Molecule type: DNA
A:Residues: 1-388 <SO6>
A:Cross-references: GB:J00279
R:Althada, S.B.P.; Tanji, M.; Kageyama, T.; Takahashi, K.
J. Biochem. 106, 920-927, 1989
A>Title: A comparative study on the NH2-terminal amino acid sequences and some other pro
A:Reference number: PX0023; MUID:90130402
A:Accession: PX0023
A:Molecule type: protein
A:Residues: 16-100 <ATH>
R:Foltmann, B.
FEBS Lett. 241, 69-72, 1988
A>Title: Activation of human pepsinogens.
A:Reference number: S02663; MUID:89065108
A:Accession: S02663
A:Molecule type: protein
A:Residues: 16-68 <FOL>
R:Ichihara, Y.; Sogawa, K.; Takahashi, K.
J. Biochem. 98, 483-492, 1985
A>Title: Isolation of human, swine, and rat prepepsinogens and calf preprochymosin, and
A:Reference number: A22434; MUID:86059312
A:Accession: F22434
A:Molecule type: protein
A:Residues: 1-15,'XXX',19-20,'X',22,'XX',25-26,'X',28 <ICH>
R:Evvers, M.P.J.; Zella, B.; Peeper, D.S.; Mager, W.H.; Planta, R.J.; Eriksson, A.W.; Fri
Hum. Genet. 77, 182-187, 1987
A>Title: Molecular cloning of a pair of human pepsinogen A genes which differ by a Glu->
A:Reference number: I54252; MUID:88006181
A:Accession: I54252
A>Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-21,'F',29-73 <RES>
A:Cross-references: GB:M27598; NID:g189834; PIDN:AAA36431.1; PID:g189836
C:Genetics:
A:Gene: GDB:PGA3
A:Cross-references: GDB:119482; OMIM:169710
A:Map position: 11q13.1-11q13.5
A:Introns: 19/2,73/3; 113/1; 152/3; 219/2; 258/2; 306/3; 339/3
C:Superfamily: pepsin
C:Keywords: aspartic proteinase; gastric juice; hydrolase; phosphoprotein; protein diges
F.1-15/Domain: signal sequence #status experimental <SIG>
F.16-388/Product: pepsinogen A 3 #status experimental <ZYM>
F.16-62/Domain: activation peptide #status experimental <APT>
F.60-388/Product: pepsin A 3, minor variant #status experimental <MIN>
F.63-388/Product: pepsin A 3 #status experimental <MAT>
F.94,277/Active site: Asp #status predicted
F.130/Binding site: phosphate (Ser) (covalent) #status predicted

Query Match          52.4%; Score 1082.5; DB 1; Length 388;
Best Local Similarity 56.7%; Pred. No. 8,2e-78;
Matches 208; Conservative 52; Mismatches 98; Indels 9; Gaps 3;

30 RLPVTKGSLRKALKRHGLLEDFLOKQOYGISKSGFEVAVS---PLTVYLDISOYRG 85
::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
19 KVPILRKSLKRLTSLRGILKDFLKNHNLNPAKKTFPPQEAFLVDEQPLIDNKEFTFG 78

```

```

QY      86  KIVLTCTPOEPTFLVDFDGGSDPWFVNSITCKSNACKNKHQREDPKRSYFUMJLGRKLSLHYG  145
      1  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      79  TIGIGTFAQDPTFVEDTGGSNLMWPVSVCSSLACTNNHNFREDSDSTYTQSTSEVSIITYG  138

QY      146  TGSMOGLIGDTVTYSNIVDIQIQVVGSLTORPGDVFITYAEFDGILGMAYPSLASEYSIPV  205
      1  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      139  TGSMTGILIGDPTVOVGGISDPTNQJFGLSETEPBGFSFLVYAPADGILGLAIPSISSGATPV  198

QY      206  FDNMMNRHLVADLFESVYMDNRNGOE-SMLTLGALDPSYYTGSLLHWVPVTVQOYMQFTVDS  264
      1  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      199  FDNIMNGLVQSODLFESVYLSLADDSGSAVITIGGIDSSVYTGSLMWVPEVYVGCYMQIIVDS  258

QY      265  VTISGVVAEGCGCCALLIDTGTSKLVGPPSSDILNIQQAIGATONQYGEFIDICDNLSTMP  324
      1  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      259  ITMNEAIALACGACGCAIYDVTGSLTGTSPSIANIQSDIGASNSDSDMVAVSCSAISLSP  318

QY      325  TVYFEINRKMPPLTPSATSDQDGFCTSGRPSSEN----HSGKMLIGDPIETREYYSVSDRA  380
      1  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      319  DIVEFTINGVQVPPPSAVTIQSEGSCTISGFQGMNLPTESGELWILGDVFIKQVFTVEDRA  378

QY      381  NNVLGIA 387
      1  : : : : :
Db      379  NNQVGLA 385

RESULT  12
B30142
pepsin A (EC 3.4.23.1) 4 precursor - human
C:Species: Homo sapiens (man)
C:Date: 07-Jun-1990 #sequence_revision 20-Aug-1994 #text_change 29-Aug-1997
C:Accession: B30142; E22434
R:Evvers, M.P.J.; Zelle, B.; Bebelman, J.P.; van Bensechem, V.; Kraakman, L.; Hoffer,
Genomics 4, 232-239, 1989
A:Title: Nucleotide sequence comparison of five human pepsinogen A (Pga) genes: evolu
A:Reference number: A91627; MUID:89233110
A:Accession: B30142
A:Molecule type: DNA
A:Residues: 1-77, 'F', 29-388 <EVE>
A:Note: the authors translated the codon TTC for residue 28 as Leu, GGC for residue 3
R:Ichihara, Y.; Sogawa, K.; Takahashi, K.
J. Biochem. 98, 483-492, 1985
A:Title: Isolation of human, swine, and rat prepepsinogens and calf picrochymosin, a
A:Reference number: A22434; MUID:86059312
A:Accession: E22434
A:Molecule type: protein
A:Residues: 1-15, 'XXX', 19-20, 'X', 22, 'XX', 25-26, 'X', 28 <ICH>
C:Genetics:
A:Gene: GDB:PGMA
A:Cross-references: GDB:119483; OMTM:169720
A:Map position: 11q13-11q13
A:introns: 19/1; 73/3; 113/1; 152/3; 219/2; 258/2; 306/3; 339/3
C:Superfamily: pepsin
C:Keywords: aspartic proteinase; hydrolase; phosphoprotein; protein digestion; zymoge
F:1-15/Domain: signal sequence #status experimental <SIG>
F:16-59/Domain: activation peptide #status experimental <APT>
F:63-388/Product: pepsin A #status predicted <MAT>
F:94, 277/Active site: Asp #status predicted
F:107-112, 268-272, 311-344/Disulfide bonds: #status predicted
F:130/Binding site: phosphate (Ser) (covalent) #status predicted

Query Match 52.3%; Score 1080.5; DB 2; Length 388;
Best Local Similarity 56.6%; Pred. No. 1.2e-77;
Matches 209; Conservative 53; Mismatches 94; Indels 13; Gaps 4;

QY      30  RIPIYKGSLLRKALKEHGLLEDFLOKQOYGISKSKSGCEGVASV-----PLNNYLDQVREG  85
      1  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      19  KVPLLRKRLSLRKLTERGGLKDFLKRKHNLPARKKFPQWEAPTLVDEQPLEYNTDMEYFG  78

QY      86  KIVYIGPPEPFVFLPTGSSDWFVPSIYCKSNACKNKHQREDFDKRSSTFONLGRKPLSIHYG  145
      1  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      79  TIGIGTFAQDPTFVEDTGGSNLMWPVSVCSSLACTNNHNFREDSDSTYTQSTSEVSIITYG  138

```







**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 31, 2002, 09:21:18 ; Search time 13.41 Seconds

(without alignments)  
1126.072 Million cell updates/sec

Title: US-09-643-755b-2

Perfect score: 2064  
Sequence: 1 MNFLKSPFPFAFLCFCFGYFY.....REYYSVDRANNLVGLAKAI 390

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt.40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1928	93.4	381	1	CHYM_BOVIN
2	1829	88.6	381	1	CHYM_SHEEP
3	1605	77.8	381	1	CHYM_CALJA
4	1108.5	53.7	387	1	PEPA_CALJA
5	1094.5	53.0	383	1	PEPE_CHICK
6	1093.5	52.9	388	1	PEP2_MACRU
7	1085.5	52.6	388	1	PEP2_MACRU
8	1084.5	52.5	388	1	PEP1_MACRU
9	1082.5	52.4	388	1	PEPA_HUMAN
10	1075.5	52.1	388	1	PEP4_MACRU
11	1070	51.8	386	1	PEPA_PIG
12	1045	50.6	387	1	PEP1_RABIT
13	1037	50.2	387	1	PEP4_RABIT
14	1034	50.1	387	1	PEP3_RABIT
15	1029	49.9	387	1	PEP2_RABIT
16	1012.5	49.1	367	1	PEPA_CHICK
17	997	48.3	388	1	PAG_HORSE
18	946	45.8	388	1	PEP1_RABIT
19	930	45.1	396	1	CATE_HUMAN
20	923	44.7	396	1	CATE_RAT
21	918	44.5	398	1	CATE_MOUSE
22	913	44.2	397	1	CATE_MOUSE
23	911.5	44.2	391	1	CATE_MOUSE
24	892	43.2	392	1	PEPC_CAVO
25	876	42.4	377	1	PEPC_RAT
26	869.5	42.1	377	1	PEPC_MACRU
27	869	42.0	388	1	PEPC_HUMAN
28	866	42.0	394	1	PEPC_CAVO
29	862	41.8	388	1	PEPC_CALJA
30	859	41.6	376	1	PAG2_BOVIN
31	807	39.1	420	1	CATD_PIG
32	797.5	38.6	412	1	CATD_CHICK
33	792.5	38.4	412	1	CATD_HUMAN

## ALIGNMENTS

34	784.5	38.0	410	1	CATD_MOUSE
35	775	37.5	407	1	CATD_RAT
36	772.5	37.4	390	1	CATD_BOVIN
37	747.5	36.2	389	1	PAG1_PIG
38	744	36.0	380	1	PAG1_BOVIN
39	728	35.3	382	1	PAG1_SHEEP
40	718	34.8	345	1	CATD_PIG
41	699	33.9	387	1	ASRP_AEDAE
42	691.5	33.5	365	1	CATD_SHEEP
43	675.5	32.7	402	1	RENI_RAT
44	670.5	32.5	400	1	RENI_SHEEP
45	668	32.4	419	1	KDAP_MOUSE

  

RESULT 1					
CHYM_BOVIN	STANDARD:	PRT:	381 AA.		
AC	P00794:				
DT	21-JUL-1986 (Rel. 01, Created)				
DT	01-JAN-1990 (Rel. 13, Last sequence update)				
DT	16-OCT-2001 (Rel. 40, Last annotation update)				
DE	Chymosin precursor (EC 3.4.23.4) (Preprorennin).				
GN	CYM OR CPC.				
OS	Bos taurus (Bovine).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;				
OX	Bovidae; Bovinae; Bos.				
OX	NCBI_TaxID=9913;				
RP	SEQUENCE FROM N.A. (CHYMOSIN B).				
RP	MEDLINE=86301873; PubMed=3091454;				
RA	Hidaka M., Sasaki K., Uozumi T., Beppu T.;				
RT	"Cloning and structural analysis of the calf prochymosin gene.;"				
RL	Gene 43:197-203(1986).				
RN	[2]				
RP	SEQUENCE FROM N.A. (CHYMOSIN B).				
RP	MEDLINE=82221400; PubMed=6283469;				
RA	Harris T.J.R., Lowe P.A., Lyons A., Thomas P.G., Eaton M.A.W.,				
RT	Millican T.A., Patel T.P., Bose C.C., Carey N.H., Doel M.T.;				
RL	"Molecular cloning and nucleotide sequence of cDNA coding for calf prochymosin.;"				
RL	Nucleic Acids Res. 10:2177-2187(1982).				
RN	[3]				
RP	SEQUENCE OF 59-381 (CHYMOSIN B), AND DISULFIDE BONDS.				
RP	MEDLINE=79239460; PubMed=381305;				
RA	Foltmann B., Pedersen V.B., Kauffman D., Wybrandt G.;				
RT	"The primary structure of calf chymosin.;"				
RL	J. Biol. Chem. 254:8447-8456(1979).				
RN	[4]				
RP	SEQUENCE OF 17-77 (CHYMOSIN B).				
RP	MEDLINE=7602411; PubMed=240697;				
RA	Pedersen V.B., Foltmann B.;				
RT	"Amino-acid sequence of the peptide segment liberated during activation of prochymosin (prorennin).;"				
RL	Eur. J. Biochem. 55:95-103(1975).				
RN	[5]				
RP	ACTIVE SITE PEPTIDES OF CHYMOSIN B.				
RP	MEDLINE=75060332; PubMed=4612029;				
RA	Chang W.-J., Takahashi K.;				
RT	"The structure and function of acid proteases. III. Isolation and characterization of the active-site peptides from bovine rennin.;"				
RL	J. Biochem. 76:467-474(1974).				
RN	[6]				
RP	SEQUENCE FROM N.A. (CHYMOSIN A).				
RP	MEDLINE=83054629; PubMed=6183168;				
RA	Moir D., Mao J., Schumm J.W., Vovis G.F., Alford B.L.,				
RT	Taunton-Rigby A.;				
RA	"Molecular cloning and characterization of double-stranded cDNA coding for bovine chymosin.;"				
RT	Gene 19:127-138(1982).				

RN [7]  
RX X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).  
RA MEDLINE-92046065; PubMed-1942052;  
RA Newman M., Saito M., Frazao C., Kahn G., Zdanov A., Tickle I.J.,  
RA Blundell T.L., Andreeva N.;  
RT "X-ray analyses of aspartic proteinases. IV. Structure and refinement  
RT at 2.2-A resolution of bovine chymosin.";  
RL J. Mol. Biol. 221:1295-1309(1991).  
RN [8]  
RX X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF MUTANT.  
RA MEDLINE-91104895; PubMed-2271625;  
RA Strop P., Sedlacek J., Stys J., Kaderabkova Z., Blaha I.,  
RA Pavlickova L., Pohl J., Fabry M., Kostka V., Newman M., Frazao C.,  
RA Shearer A., Tickle I.J., Blundell T.L.;  
RT "Engineering enzyme substrate specificity: preparation, kinetic  
RT characterization, and X-ray analysis at 2.0-A resolution of Val111Phe  
RT site-mutated calf chymosin.";  
RL Biochemistry 29:9863-9871(1990).  
RN [9]  
RX X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).  
RA MEDLINE-91017501; PubMed-2217166;  
RA Gilliland G.L., Winborne E.L., Nachman J., Wlodawer A.;  
RT "The three-dimensional structure of recombinant bovine chymosin at  
RT 2.3-A resolution.";  
RL Proteins 8:82-101(1990).  
RN [10]  
RP MUTAGENESIS OF CYS-308 AND CYS-341.  
RX MEDLINE-92412108; PubMed-1530626;  
RX Huang K., Zhang Z., Liu N., Zhang Y., Zhang G., Yang K.;  
RT "Functional implication of disulfide bond, Cys250-Cys283, in bovine  
RT chymosin.";  
RL Biochem. Biophys. Res. Commun. 187:692-696(1992).  
CC -1- FUNCTION: CHYMOSIN IS SYNTHESIZED IN THE MUCOSA OF THE ABOMASUM  
CC (FOURTH STOMACH) OF YOUNG (UNWEANED) RUMINANTS. THE ENZYME  
CC HYDROLYSES CASEIN TO PARACASEIN.  
CC -1- CATALYTIC ACTIVITY: Broad specificity similar to that of pepsin A.  
CC Clots milk by cleavage of a single bond in casein (kappa chain).  
CC -1- SUBUNIT: MONOMER.  
CC -1- POLYMORPHISM: FORMS A AND B ARE PROBABLY ALLELIC VARIANTS. FORM B  
CC IS THE PREDOMINANT FORM AND IS THE SEQUENCE SHOWN.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE  
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: M14077; AAA30446.1; -;  
DR EMBL: M14069; AAA30446.1; JOINED.  
DR EMBL: M14070; AAA30446.1; JOINED.  
DR EMBL: M14071; AAA30446.1; JOINED.  
DR EMBL: M14072; AAA30446.1; JOINED.  
DR EMBL: M14073; AAA30446.1; JOINED.  
DR EMBL: M14074; AAA30446.1; JOINED.  
DR EMBL: M14075; AAA30446.1; JOINED.  
DR EMBL: J00002; AAA30447.1; -;  
DR EMBL: J00003; AAA30448.1; -;  
DR PIR: A00985; CMOB.  
DR PIR: A25631; A25631.  
DR PDB: 1CMS; 15-JUL-92.  
DR PDB: 3CMS; 15-OCT-92.  
DR PDB: 4CMS; 15-OCT-91.  
DR PDB: 1CZ1; 01-APR-97.  
DR MEROPS; A01.006; -;  
DR InterPro: IPR001969; Asp-protease.  
DR InterPro: IPR001461; Pepsin.  
DR Pfam: PF00026; asp. 1  
DR PRINTS; PR00792; PEPsin.  
DR PROSITE; PS00141; ASP\_PROTEASE; 2.

KW Hydrolase; Aspartyl protease; Digestion; Stomach; Gastric juice;  
KW Zymogen; Signal; 3D-structure.  
FT SIGNAL 1 16  
FT PROPEP 17 58  
FT CHAIN 59 381  
FT ACT\_SITE 92 92  
FT ACT\_SITE 274 274  
FT DISULFID 105 110  
FT DISULFID 265 269  
FT DISULFID 308 341  
FT VARIANT 302 302  
FT CONFLICT 17 109  
FT CONFLICT 109 109  
FT CONFLICT 155 155  
FT CONFLICT 218 218  
FT CONFLICT 230 230  
FT CONFLICT 232 232  
FT STRAND 62 69  
FT STRAND 70 72  
FT STRAND 73 80  
FT STRAND 81 84  
FT STRAND 85 92  
FT TURN 93 94  
FT STRAND 98 102  
FT TURN 103 104  
FT HELIX 108 111  
FT TURN 112 112  
FT STRAND 116 116  
FT HELIX 118 120  
FT TURN 122 123  
FT STRAND 125 135  
FT TURN 136 137  
FT STRAND 138 151  
FT TURN 152 153  
FT STRAND 154 166  
FT HELIX 170 174  
FT STRAND 179 182  
FT HELIX 186 188  
FT TURN 191 192  
FT TURN 196 202  
FT HELIX 203 204  
FT STRAND 210 214  
FT TURN 217 218  
FT STRAND 222 226  
FT HELIX 231 233  
FT STRAND 234 242  
FT STRAND 246 246  
FT TURN 247 248  
FT STRAND 249 258  
FT TURN 259 260  
FT STRAND 261 264  
FT TURN 266 267  
FT STRAND 269 273  
FT TURN 275 276  
FT STRAND 280 282  
FT HELIX 284 294  
FT TURN 295 295  
FT STRAND 297 299  
FT TURN 300 302  
FT STRAND 303 306  
FT TURN 308 309  
FT TURN 311 313  
FT STRAND 317 321  
FT TURN 322 323  
FT STRAND 324 328  
FT HELIX 330 333  
FT STRAND 334 337  
FT TURN 338 339  
FT STRAND 340 343  
FT STRAND 345 347  
FT STRAND 354 356  
FT HELIX 358 361  
FT TURN 362 363



Query Match 93.4%; Score 1928; DB 1; Length 381;  
Best Local Similarity 98.6%; Pred. No. 5,2e-140;  
Matches 365; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 21 AVTHAAETRIPLKYGKSLRKALKEHGLEDFLOKQOYGSSKYSGFGEVASVPLNTYLD 80  
::: |||||  
DB 12 ALSQGAETIRIPLKYGKSLRKALKEHGLEDFLOKQOYGSSKYSGFGEVASVPLNTYLD 71  
QY 81 SOYFGKITLGPPOEFTVLFDTGSSDFWVPSTYCKSNACKNQRDPKRSSTFQNLGKPL 140  
||| |||||  
DB 72 SOYFGKITLGPPOEFTVLFDTGSSDFWVPSTYCKSNACKNQRDPKRSSTFQNLGKPL 131  
QY 141 SIHGTGSMQGLIGDYDTVYNSIVDIQOTVGLSTQEPGDVFTYAEFDGILGMAVPSLASE 200  
||| |||||  
DB 132 SIHGTGSMQGLIGDYDTVYNSIVDIQOTVGLSTQEPGDVFTYAEFDGILGMAVPSLASE 191  
QY 201 YSIPFEDNMNRHLVAODLFSVYMDRNGQESMLTLGAIDPSYTGSLHWVPYVQOYWOE 260  
||| |||||  
DB 192 YSIPFEDNMNRHLVAODLFSVYMDRNGQESMLTLGAIDPSYTGSLHWVPYVQOYWOE 251  
QY 261 TVDSVTISGVVACGCGQALIDTGTSKLVGPSSDILNIQOALIGATONQYGFEDIDCNL 320  
||| |||||  
DB 252 TVDSVTISGVVACGCGQALIDTGTSKLVGPSSDILNIQOALIGATONQYGFEDIDCNL 311  
QY 321 SYMPVTVFEINCKMPLTPSATYSODGFCSTGFOSENHSHQMLIGDVFIREYYSVPDRA 380  
||| |||||  
DB 312 SYMPVTVFEINCKMPLTPSATYSODGFCSTGFOSENHSHQMLIGDVFIREYYSVPDRA 371  
QY 381 NNLVGLAKAI 390  
||| |||||  
DB 372 NNLVGLAKAI 381

RESULT 2  
CHYM\_SHEEP STANDARD; PRT; 381 AA.  
AC P18276;  
DT 01-NOV-1990 (Rel. 16, Created)  
DT 01-NOV-1990 (Rel. 16, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Chymosin precursor (EC 3.4.23.4) (Preprorennin).  
GN CVM.  
OS Ovis aries (Sheep).  
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Caprinae; Ovis.  
OX NCBI\_TaxID=9940;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90356410; PubMed=2117748;  
RA Pungekar J., Strukelj B., Gubensek F., Turk V., Kregar I.;  
RT "Complete primary structure of lamb preprochymosin deduced from  
RT cDNA".  
RL Nucleic Acids Res. 18:4602-4602(1990).  
CC -!- FUNCTION: CHYMOSIN IS SYNTHESIZED IN THE MUCOSA OF THE STOMACH.  
CC THE ENZYME HYDOLYZES CASEIN TO PARACASEIN.  
CC -!- CATALYTIC ACTIVITY: Broad specificity similar to that of pepsin A.  
CC Cuts milk by cleavage of a single bond in casein (kappa chain).  
CC -!- SUBUNIT: MONOMER.  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE  
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.  
CC -----  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X53037; CAJ37209.1; -;  
DR PIR; S10996; CMSHB.

DR HSSP; P00794; ICMS.  
DR MEROPS; A01.006; -;  
DR InterPro; IPR001969; Asp\_protease.  
DR InterPro; IPR001461; Pepsin.  
DR Pfam; PF00026; asp. 1.  
DR PRINTS; PR00792; PEPsin.  
DR PROSITE; PS00141; ASP\_PROTEASE; 2.  
KW Hydrolase; Aspartyl protease; Digestion; Stomach; Gastric juice;  
KW zymogen; Signal.  
FT SIGNAL 1 16  
FT PROPEP 17 58  
FT CHAIN 59 381  
FT ACT\_SITE 92 92  
FT ACT\_SITE 274 274  
FT DISULFID 105 110  
FT DISULFID 265 269  
FT DISULFID 308 341  
SQ SEQUENCE 381 AA; 42074 MW; D9903528FA071C47 CRC64;

Query Match 88.6%; Score 1829; DB 1; Length 381;  
Best Local Similarity 93.2%; Pred. No. 1.9e-132;  
Matches 345; Conservative 13; Mismatches 12; Indels 0; Gaps 0;

QY 21 AVTHAAETRIPLKYGKSLRKALKEHGLEDFLOKQOYGSSKYSGFGEVASVPLNTYLD 80  
::: |||||  
DB 12 ALSQGAETIRIPLKYGKSLRKALKEHGLEDFLOKQOYGSSKYSGFGEVASVPLNTYLD 71  
QY 81 SOYFGKITLGPPOEFTVLFDTGSSDFWVPSTYCKSNACKNQRDPKRSSTFQNLGKPL 140  
||| |||||  
DB 72 SOYFGKITLGPPOEFTVLFDTGSSDFWVPSTYCKSNACKNQRDPKRSSTFQNLGKPL 131  
QY 141 SIHGTGSMQGLIGDYDTVYNSIVDIQOTVGLSTQEPGDVFTYAEFDGILGMAVPSLASE 200  
||| |||||  
DB 132 SIHGTGSMQGLIGDYDTVYNSIVDIQOTVGLSTQEPGDVFTYAEFDGILGMAVPSLASE 191  
QY 201 YSIPFEDNMNRHLVAODLFSVYMDRNGQESMLTLGAIDPSYTGSLHWVPYVQOYWOE 260  
||| |||||  
DB 192 YSIPFEDNMNRHLVAODLFSVYMDRNGQESMLTLGAIDPSYTGSLHWVPYVQOYWOE 251  
QY 261 TVDSVTISGVVACGCGQALIDTGTSKLVGPSSDILNIQOALIGATONQYGFEDIDCNL 320  
||| |||||  
DB 252 TVDSVTISGVVACGCGQALIDTGTSKLVGPSSDILNIQOALIGATONQYGFEDIDCNL 311  
QY 321 SYMPVTVFEINCKMPLTPSATYSODGFCSTGFOSENHSHQMLIGDVFIREYYSVPDRA 380  
||| |||||  
DB 312 SYMPVTVFEINCKMPLTPSATYSODGFCSTGFOSENHSHQMLIGDVFIREYYSVPDRA 371  
QY 381 NNLVGLAKAI 390  
||| |||||  
DB 372 NNLVGLAKAI 381

RESULT 3  
CHYM\_CALJA STANDARD; PRT; 381 AA.  
AC Q9N2D2;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Chymosin precursor (EC 3.4.23.4) (Preprorennin).  
GN CVM.  
OS Callitrix jacchus (Common marmoset).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Platyrrhini; Callitrichidae;  
OC Callitrich.  
OX NCBI\_TaxID=9483;  
RN [1]  
RP SEQUENCE FROM N.A.; SEQUENCE OF 17-26, FUNCTION, AND ENZYME  
RP REGULATION.  
RC TISSUE=Gastric mucosa;  
RX MEDLINE=20250834; PubMed=10788784;  
RA Kageyama T.;

```

RT "New World monkey pepsinogens A and C, and prochymosins. Purification,
RT characterization of enzymatic properties, cDNA cloning, and molecular
RT evolution."
RT
CC J. Biochem. 127:761-770(2000).
CC
CC -1- FUNCTION: Hydrolyses a variety of proteins.
CC -1- CATALYTIC ACTIVITY: Broad specificity similar to that of pepsin A.
CC Clots milk by cleavage of a single bond in casein (kappa chain).
CC -1- ENZYME REGULATION: Inhibited by pepstatin.
CC -1- SUBUNIT: Monomer (By similarity).
CC -1- DEVELOPMENTAL STAGE: Expressed in adult, not neonate-specific as
CC in other organisms.
CC -1- MISCELLANEOUS: The optimal pH is around 2.5.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC
CC EMBL; AB038386; BAA90873.1; -.
CC InterPro; IPR001969; Asp-protease.
CC DR Pfam; PF000026; asp. 1.
CC DR PRINTS; PR00792; PEPsin.
CC DR PROSITE; PS00141; ASP-PROTEASE; 2.
CC KM Hydrolyase; Aspartyl protease; Digestion; Gastric juice; Zymogen;
CC Repeat; Signal.
CC FT SIGNAL 1 16
CC FT PROPEP 17 58 ACTIVATION PEPTIDE (BY SIMILARITY).
CC FT CHAIN 59 381 CHYMOSIN.
CC FT REPEAT 92 102 1.
CC FT REPEAT 274 284 2.
CC FT ACT_SITE 92 92 BY SIMILARITY.
CC FT ACT_SITE 274 274 BY SIMILARITY.
CC FT DISULFID 105 110 BY SIMILARITY.
CC FT DISULFID 265 269 BY SIMILARITY.
CC FT DISULFID 308 341 BY SIMILARITY.
CC SQ SEQUENCE 381 AA; 41896 MW; C5820C74C97BB96B CRC64;

Query Match 77.8%; Score 1605; DB 1; Length 381;
Best Local Similarity 81.4%; Pred. No. 2,4e-115;
Matches 301; Conservative 30; Mismatches 39; Indels 0; Gaps 0;

```

```

Db 372 SNLVGLAKAI 381
:|||||
RESULT 4
PEPA_CALJA STANDARD; PRT; 387 AA.
ID PEPA_CALJA
AC O9N2D4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Pepsin A precursor (EC 3.4.23.1).
GN PEA.
OS Callithrix jacchus (Common marmoset).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Platyrrhini; Callitrichidae;
OC Callithrix.
OX NCBI_TaxID=9483;
RN (1)
RP SEQUENCE FROM N.A., SEQUENCE OF 16-25, FUNCTION, AND ENZYME
RP REGULATION.
RP TISSUE=Gastric mucosa;
RX MEDLINE=20250834; Pubmed=10788784;
RA Kageyama T.;
RT "New World monkey pepsinogens A and C, and prochymosins. Purification,
RT characterization of enzymatic properties, cDNA cloning, and molecular
RT evolution."
RT
CC J. Biochem. 127:761-770(2000).
CC
CC -1- FUNCTION: Shows particularly broad specificity; although bonds
CC involving phenylalanine and leucine are preferred, many others are
CC also cleaved to some extent.
CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE PHE-, LEU-.
CC -1- ENZYME REGULATION: Inhibited by pepstatin.
CC -1- MISCELLANEOUS: The optimal pH is around 2.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC
CC EMBL; AB038384; BAA90871.1; -.
CC InterPro; IPR001969; Asp-protease.
CC DR Pfam; PF000026; asp. 1.
CC DR PRINTS; PR00792; PEPsin.
CC DR PROSITE; PS00141; ASP-PROTEASE; 2.
CC KM Hydrolyase; Aspartyl protease; Digestion; Gastric juice; Zymogen;
CC Signal; Phosphorylation.
CC FT SIGNAL 1 15
CC FT PROPEP 16 61 ACTIVATION PEPTIDE (BY SIMILARITY).
CC FT CHAIN 62 387 PEPsin A.
CC FT MOD_RES 129 129 PHOSPHORYLATION (BY SIMILARITY).
CC FT ACT_SITE 93 93 BY SIMILARITY.
CC FT ACT_SITE 276 276 BY SIMILARITY.
CC FT DISULFID 106 111 BY SIMILARITY.
CC FT DISULFID 267 271 BY SIMILARITY.
CC FT DISULFID 310 343 BY SIMILARITY.
CC SQ SEQUENCE 387 AA; 41563 MW; 7A7968A568464BD CRC64;

Query Match 53.7%; Score 1108.5; DB 1; Length 387;
Best Local Similarity 58.6%; Pred. No. 1.9e-77;
Matches 215; Conservative 49; Mismatches 94; Indels 9; Gaps 3;

```

OY	86	KIYLGTPOEETVLFEDRGSSDFWVPSTLYCKSNMKNORDPKRSPFOULGRLPIHNG	145
Dd	78	TIGTGPQOETVTFTFDGSSNLWWPSTLYCSSPATNNRRNPQSSSYQATSOFLSLTG	1377
OY	146	TGSNQGILGYDTVTVNSNIVDIQOTVGGLSTEPGDVFTEFAEFGDILGAAYSLASESIPV	2051
Dd	138	TGSMWTGILGYDTVGVGGIADTNQIFGLSETFPGSFLLYSFPGDILGLAIATSISSGATPV	197
OY	206	FDDNMNRHLVAQDLFSYYMDRNQE-SMLTLGAIDPSYTYTSGSLHWMPVTVQQXYMOFTVDS	264
Dd	198	FDNIHQDVLVSQDLFSYYLSSNDQSGSVMEFGIDSSYTGSLMWVPVSAEGTWQITVDS	257
OY	265	VTISGVVAVCEGCQAILDFTGTSKLVGPSDDILNIQQAIGATONQXGEFDIDCDNLSTYP	324
Dd	258	ITTMGEALICAECCQAIVDTGTSLLSGTPSPIANIGCYIGASENSNGEMVSCSAISLP	317
OY	325	TVVEINGMKMTPLTPSAVTSODGFCISGROSEN---HSQAKILGDVFTREYTSVFDRA	360
Dd	318	DIVETINGIOVPPASAVYIIODECGCTSGOGMNIPRAYELMTILGCVFIROYFAVEDRA	377
OY	381	NNLVGLA 387	
Dd	378	NNOVGLA 384	

PEPE_RESULT	5			
PEPE_CHICK	ID	PEPE_CHICK	STANDARD:	PRT: 383 AA.
AC	P16476;			
DT	01-AUG-1990 (Rel. 15, Created)			
DT	01-AUG-1990 (Rel. 15, Last sequence update)			
DT	01-NOV-1995 (Rel. 32, Last annotation update)			
DE	Embryonic pepsinogen precursor (EC 3.4.23.-).			
OS	Gallus gallus (chicken).			
OC	Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi;			
OC	Archosauromia; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;			
OC	Gallus.			
OX	NCBI_TaxID=9031;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=688227903; PubMed=313317;			
RA	Hayashi K., Agata K., Mochii M., Yasugi S., Eguchi G., Mizuno T.;			
RT	Molecular cloning and the nucleotide sequence of cDNA for embryonic			
RT	chicken pepsinogen: phylogenetic relationship with prochymosin.;			
RL	J. Biochem. 103:290-296(1988).			
CC	-1- DEVELOPMENTAL STAGE: SPECIFICALLY SECRETED DURING THE EMBRYONIC			
CC	PERIOD IN THE CHICKEN PROVENTRICULUS (GLANDULAR STOMACH).			
CC	-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE			
CC	EUKARYOTIC ASPARTYL PROTEASES FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
DR	EMBL, D00215; BAA00153.1; -.			
DR	PIR; A41443; A41443.			
DR	HSP; P00794; ACMS.			
DR	MEROPS; A01.028; -.			
DR	InterPro: IPR001969; Asp.protease.			
DR	InterPro: IPR001461; Pepsin.			
DR	Pfam: PF00026; asp.1.			
DR	PRINTS; PR00792; PEPsin.			
DR	PROSITE; PS00141; ASP_PROTEASE; 2.			
KW	Hydrolase; Aspartyl protease; Digestion; Signal; Glycoprotein.			
FT	CHAIN 1 383			
FT	ACT_SITE 94 94			
FT	ACT_SITE 276 276			
FT	DISULFID 107 112			
FT				
FT	POTENTIAL.			
FT	EMBRYONIC PEPSINOGEN.			
FT	BY SIMILARITY.			
FT	BY SIMILARITY.			
FT	BY SIMILARITY.			

FT	DISULFID	267	271	BY SIMILARITY.
FT	CARBOHYD	310	344	BY SIMILARITY.
FT	CARBOHYD	132	132	N-LINKED (GLCNAC. . .) (POTENTIAL)
FT	CARBOHYD	204	204	N-LINKED (GLCNAC. . .) (POTENTIAL)
FT	CARBOHYD	309	309	N-LINKED (GLCNAC. . .) (POTENTIAL)
FT	CARBOHYD	350	350	N-LINKED (GLCNAC. . .) (POTENTIAL)
FT	VARIANT	51	51	T -> S.
Q	SEQUENCE	383 AA;	41719 MW;	1642796871611P54 CRC64;

Query Match	53.0%;	Score 1094.5;	DB 1;	Length 383;
Best Local Similarity	-54.4%;	Pred. No. 2.2e-76;		
Matches 210; Conservative	68;	Mismatches 95;	Indels 13;	Gaps 6;

QY	11	AFLEGGYFAVAYNAAE--IYRIPRYKRSKRLAKREHLEDEFLQKOOYISRRY--SGFG	68
Db	5	ALLC-----AVALLASGITRRLPERLSEKKLREILIREKGLHNHFLOHNRYDIGKRFNAPR	58
QY	69	EYASV---PLTNALDISOYFECKIYLGPRIPEFYVLFEDTGSDDFVWPVSYCKSNACKNHQRE	125
Db	59	DVLVLEPEPLNLTLDMEYTGISTIGRPDDEFTVWFEDTGSNNLWVPVSYCTSPACSHOME	118
QY	126	DPKRSTFEONLGRKPLSHVYGTSGMOGILGVDYTVASNIYVIOQTGVLSTQEPEDVDFTYAE	185
Db	119	NPSSSTSTYKSTGQNLSTHYGTSGDMEGTGVGDYTVASLMTNOLFGLSTSEPOQFVYUAK	178
QY	186	FDGLIGAAVPSLASEYSIPEDNNMMNHLVAOOLFESYMDNRQOESMLTGLAIDPSPYTG	245
Db	179	FDGLIGLIGTYSLAADGTTPEVFDNNWNESLLEONLFVSYLSREPMGSMVYFGGIDESTFTG	238
QY	246	SLHWVPYVTOOYMOQFTVDVSYTISGVVVACEGCGOAILDSTSKLVGPSSDILNIQOALGA	305
Db	239	SIMNIPEVSYGQWQMSDLSITVKKQELACSSGOAIDTGTSLVAGPAAINDIQSAVGA	298
QY	306	TQNOYGEFDICDNLSTMPRVVEIINGKATPLTPSAVTSOD--QGFTSGEQSESNHSQAKI	364
Db	299	NQNTYGEYSYNCSHILAMPVVEVFIIGIQYPRPALAYTEQNGOGCTMSSPONS--ADLWT	357
QY	365	LGDVEIREYYSVFDRANNLIGLAKAI	390
Db	358	LGDVEIRVYISIEDRANNVGLAKAI	383

	RESULT	6
PEP2_MACFU	ID	PEP2_MACFU
	STANDARD:	PRT; 388 AA.
AC	p27677;	
DT	01-AUG-1992 (Rel. 23, Created)	
DT	01-AUG-1992 (Rel. 23, Last annotation update)	
DT	30-MAY-2000 (Rel. 39, Last annotation update)	
DE	Pepsin A-2/A-3 precursor (EC 3.4.23.1) (Pepsin III-2/III-1).	
OC	Macaca fuscata fuscata (Japanese macaque).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;	
OC	Cercopithecinae; Macaca.	
OX	NCBI_TaxId=9543;	
RN	[1]	
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 16-70.	
RC	TISSUE=Gastric mucosa;	
RX	MEDLINE=92037645; PubMed=1935977;	
RA	Kageyama T., Tanabe K., Kolwal O.;	
RT	"Development-dependent expression of isozymogens of monkey	
RL	pepsinogens and structural differences between them.";	
CC	Eur. J. Biochem. 202:205-215(1991).	
CC	-1- FUNCTION: SHOWS PARTICULARLY BROAD SPECIFICITY; ALTHOUGH BONDS	
CC	INVOLVING PHENYLALANINE AND LEUCINE ARE PREFERRED, MANY OTHERS ARE	
CC	ALSO CLEAVED TO SOME EXTENT.	
CC	-1- CATALYTIC ACTIVITY: PREPERENTIAL CLEAVAGE PHE-, LEU-.	
CC	-1- DEVELOPMENTAL STAGE: PEP A-2 IS DOMINANT AT THE 4-MONTH STAGE.	
CC	PEP A-3 IS DOMINANT AT FETAL STAGES.	
CC	-1- PTM: PEP A-2 IS PHOSPHORYLATED, BUT NOT PEP A-3.	
CC	-1- PTM: EACH PEPSINOGEN IS CONVERTED TO CORRESPONDING PEPSIN AT PH	
CC	2.0 IN PART AS A RESULT OF THE RELEASE OF A 47 AA ACTIVATION	

```
CC SEGMENT AND IN PART AS A RESULT OF STEPWISE PROTEOLYTIC CLEAVAGE
CC VIA AN INTERMEDIATE FORM(S).
CC -1- MISCELLANEOUS: THE EXPRESSION OF PEPSINOGEN GENES IS REGULATED BY
CC HORMONES AND RELATED SUBSTANCES.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
CC -1- CAUTION: IT IS NOT KNOWN IF THIS IS PEP A-2 OR PEP A-3.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X59755; CAA2427.1; -.
DR PIR: S16064; S16064.
DR PIR: S19684; S19684.
DR HSSP: P00790; IPSN.
DR MEROPS: A01.001; -.
DR InterPro: IPR001969; Asp.protease.
DR InterPro: IPR001791; Laminin_G.
DR Pfam: PF00026; asp. 1.
DR PRINTS: PR00792; PEPsin.
DR PROSITE: PS00141; ASP_PROTEASE; 2.
KM Hydroxylase: Aspartyl protease; Digestion: Stomach; Gastric juice;
KW Zymogen; Multigene family; Phosphorylation; Signal.
FT SIGNAL 1 15
FT PROPEP 16 40 ACTIVATION PEPTIDE.
FT PROPEP 41 62 ACTIVATION PEPTIDE.
FT CHAIN 63 388 PEPsin A-2/A-3.
FT ACT_SITE 94 94 BY SIMILARITY.
FT ACT_SITE 277 277 BY SIMILARITY.
FT DISULFID 107 112 BY SIMILARITY.
FT DISULFID 268 272 BY SIMILARITY.
FT DISULFID 311 344 BY SIMILARITY.
FT MOD_RES 130 130 PHOSPHORYLATION (POTENTIAL).
SQ SEQUENCE 388 AA; 41703 MW; 706f7ED50F59CDD CRC64;

Query Match 52.9%; Score 1092.5; DB 1; Length 388;
Best Local Similarity 56.9%; Pred. No. 3.2e-76;
Matches 210; Conservative 53; Mismatches 97; Indels 9; Gaps 3;

OY 28 ITRIPLYKGSRLKALKHGLLEDFLOKQOYGISKSYSGFEVAVS-----PLTYNLDISQY 83
DB 17 IHKVLVRRKSLRRNLSEHGLKDFLKKHNNRNPASKYFPQAEPFLIDEOPLENLDVEY 76
OY 84 FGKTYLGPPEQFVLFPTDSSDFWPSIYCKSNACKNHOFRDKRSTFQNLGKPLSIH 143
DB 77 FGTIGTGPADFTVIYIDTGSNNLWPSYVCSLACTNHNFNQDSSITYSTGTSVIT 136
OY 144 YGTSGMOGILGDTVTYVSNVLDIOQTGVLSTQEPGDVFTVAEPDGLIGMAVPSLASEYSI 203
DB 137 YGTSGMFGILGDTVTYVSGISDTNQIFGLSETEPGSFLYVAPFGIIGLAVPSISSGAT 196
OY 204 PVFENNAHRHVAADLSVYVDRNGOE-SMLTLGAIDPSYTGSLHWVRYVOQWQTV 262
DB 197 PVFNNINMGIVSODLFSVYLSADDOGSVYIFGIDSSYTGSLMWVPSVEGWQISV 256
OY 263 DSVYISGVVAVCEGGAQLIDTGTSLKLVGPSDDLINQOAGATONQVGEEDIDQDNLSY 322
DB 257 DSIIMNGEALACARCGQIVDTGSLTLGFTSPIANIOSDIGASNSGCEWVSCSAISS 316
OY 323 MPTVVEINGKMYPLTPSATYSODQFCTSGFQS-----ENHSQMLIGDVTIREYVSVD 378
DB 317 LPDVFETINGIQYVPVSATYILQSGSCISGFGMDVPTESGELWILGDVIRQYFTVFD 376
OY 379 RANNVGLA 387
DB 377 RANNOVGIA 385
```

```
RESULT 7
PEPA_MACMU STANDARD; PRT; 388 AA.
ID PEPA_MACMU
AC P11489;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Pepsin A precursor (EC 3.4.23.1).
GN pga.
OS Macaca mulatta (Rhesus macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Macaca.
OX NCBI_TaxID=9544;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88313666; PubMed=2900796;
RA Evers M.P.J., Zelle B., Bebelman J.P., Pronk J.C., Mager W.H.,
RA Planta R.J., Eriksson A.W., Frants R.R.;
RT "Cloning and sequencing of rhesus monkey pepsinogen A cDNA.";
RL Gene 65:179-185(1988).
RN [2]
RP REVISION.
RA Zelle B.;
RL Submitted (SEP-1988) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: SHOWS PARTICULARLY BROAD SPECIFICITY; ALTHOUGH BONDS
CC INVOLVING PHENYLALANINE AND LEUCINE ARE PREFERRED, MANY OTHERS ARE
CC ALSO CLEAVED TO SOME EXTENT.
CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE PHE-, LEU-.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M20788; AAA36902.1; -.
DR PIR: JT0309; PEMQAR.
DR HSSP: P00790; IPSN.
DR MEROPS: A01.001; -.
DR InterPro: IPR001969; Asp.protease.
DR InterPro: IPR001791; Laminin_G.
DR InterPro: IPR001461; Pepsin.
DR Pfam: PF00026; asp. 1.
DR PRINTS: PR00792; PEPsin.
DR PROSITE: PS00141; ASP_PROTEASE; 2.
KM Hydroxylase: Aspartyl protease; Digestion: Stomach; Gastric juice;
KW Zymogen; Signal; Phosphorylation.
FT SIGNAL 1 15
FT PROPEP 16 62 ACTIVATION PEPTIDE.
FT CHAIN 63 388 PEPsin A.
FT MOD_RES 130 130 PHOSPHORYLATION (POTENTIAL).
FT ACT_SITE 94 94
FT ACT_SITE 277 277
FT DISULFID 107 112 BY SIMILARITY.
FT DISULFID 268 272 BY SIMILARITY.
FT DISULFID 311 344 BY SIMILARITY.
SQ SEQUENCE 388 AA; 41696 MW; 97f65E3F6C2A793 CRC64;

Query Match 52.6%; Score 1085.5; DB 1; Length 388;
Best Local Similarity 57.2%; Pred. No. 1.1e-75;
Matches 211; Conservative 51; Mismatches 98; Indels 9; Gaps 3;

OY 28 ITRIPLYKGSRLKALKHGLLEDFLOKQOYGISKSYSGFEVAVS-----PLTYNLDISQY 83
DB 17 IHKVLVRRKSLRRNLSEHGLKDFLKKHNNRNPASKYFPQAEPFLIDEOPLENLDVEY 76
```



AC P00790; 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Pepsin A precursor (EC 3.4.23.1).  
GN PGAS3 AND PGAS4 AND PGAS5.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
ON NCBI\_TaxID=9606;  
RX MEDLINE=63161158; PubMed=6300126;  
RA Sogawa K., Fujii-Kuriyama Y., Mizukami Y., Ichihara Y., Takahashi K.;  
RT "Primary structure of human pepsinogen gene.";  
RL J. Biol. Chem. 258:5306-5311(1983).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOZYME 5).  
RC TISSUE=Placenta;  
RX MEDLINE=69233110; PubMed=2714789;  
RA Evers M.P.J., Zelle B., Bebelman J.P., van Beusechem V., Kraakman L.,  
RA Hoffer M.J.V., Pronk J.C., Mager W.H., Planta R.J., Eriksson A.W.,  
RA Frants R.K.;  
RT "Nucleotide sequence comparison of five human pepsinogen A (PGA)  
RT genes: evolution of the PGA multigene family.";  
RL J. Biochem. 4:232-239(1989).  
RN [3]  
RP PARTIAL SEQUENCE OF 1-28.  
RX MEDLINE=86059312; PubMed=2415509;  
RA Ichihara Y., Sogawa K., Takahashi K.;  
RT "Isolation of human, swine, and rat prepepsinogens and calf  
RT preprochymosin, and determination of the primary structures of their  
RT NH2-terminal signal sequences.";  
RL J. Biochem. 98:483-492(1985).  
RN [4]  
RP SEQUENCE OF 16-100 (ISOZYMES 2, 3; 3A; 4 AND 5).  
RX MEDLINE=90130402; PubMed=2515193;  
RA Athauda S.B.P., Tanji M., Kageyama T., Takahashi K.;  
RT "A comparative study on the NH2-terminal amino acid sequences and  
RT some other properties of six isozymic forms of human pepsinogens and  
RT pepsins.";  
RL J. Biochem. 106:920-927(1989).  
RN [5]  
RP SEQUENCE OF 16-68 (ISOZYMES 3 AND 5).  
RX MEDLINE=89065108; PubMed=3197840;  
RA Foltmann B.;  
RT "Activation of human pepsinogens.";  
RL FEBS Lett. 241:69-72(1988).  
RN [6]  
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).  
RX MEDLINE=95392399; PubMed=766352;  
RA Fujiwaga M., Chernaia M.M., Tarasova N.I., Mosimann S.C.,  
RA James M.N.G.;  
RT "Crystal structure of human pepsin and its complex with pepstatin.";  
RL Protein Sci. 4:960-972(1995).  
CC -I- FUNCTION: SHOWS PARTICULARLY BROAD SPECIFICITY, ALTHOUGH BONDS  
CC INVOLVING PHENYLALANINE AND LEUCINE ARE PREFERRED, MANY OTHERS ARE  
CC ALSO CLEAVED TO SOME EXTENT.  
CC -I- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE PHE-, LEU-  
CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE  
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL; J00287; AAA98529.1; -  
DR EMBL; J00279; AAA98529.1; JOINED.  
DR EMBL; J00280; AAA98529.1; JOINED.

DR EMBL; J00281; AAA98529.1; JOINED.  
DR EMBL; J00282; AAA98529.1; JOINED.  
DR EMBL; J00283; AAA98529.1; JOINED.  
DR EMBL; J00284; AAA98529.1; JOINED.  
DR EMBL; J00285; AAA98529.1; JOINED.  
DR EMBL; J00286; AAA98529.1; JOINED.  
DR EMBL; J00287; AAA98529.1; JOINED.  
DR EMBL; J00288; AAA98529.1; JOINED.  
DR EMBL; J00289; AAA98529.1; JOINED.  
DR EMBL; J00290; AAA98529.1; JOINED.  
DR EMBL; J00291; AAA98529.1; JOINED.  
DR EMBL; J00292; AAA98529.1; JOINED.  
DR EMBL; J00293; AAA98529.1; JOINED.  
DR EMBL; J00294; AAA98529.1; JOINED.  
DR EMBL; J00295; AAA98529.1; JOINED.  
DR EMBL; J00296; AAA98529.1; JOINED.  
DR EMBL; J00297; AAA98529.1; JOINED.  
DR EMBL; J00298; AAA98529.1; JOINED.  
DR EMBL; J00299; AAA98529.1; JOINED.  
DR EMBL; J00300; AAA98529.1; JOINED.  
DR EMBL; J00301; AAA98529.1; JOINED.  
DR EMBL; J00302; AAA98529.1; JOINED.  
DR EMBL; J00303; AAA98529.1; JOINED.  
DR EMBL; J00304; AAA98529.1; JOINED.  
DR EMBL; J00305; AAA98529.1; JOINED.  
DR EMBL; J00306; AAA98529.1; JOINED.  
DR EMBL; J00307; AAA98529.1; JOINED.  
DR EMBL; J00308; AAA98529.1; JOINED.  
DR EMBL; J00309; AAA98529.1; JOINED.  
DR EMBL; J00310; AAA98529.1; JOINED.  
DR EMBL; J00311; AAA98529.1; JOINED.  
DR EMBL; J00312; AAA98529.1; JOINED.  
DR EMBL; J00313; AAA98529.1; JOINED.  
DR EMBL; J00314; AAA98529.1; JOINED.  
DR EMBL; J00315; AAA98529.1; JOINED.  
DR EMBL; J00316; AAA98529.1; JOINED.  
DR EMBL; J00317; AAA98529.1; JOINED.  
DR EMBL; J00318; AAA98529.1; JOINED.  
DR EMBL; J00319; AAA98529.1; JOINED.  
DR EMBL; J00320; AAA98529.1; JOINED.  
DR EMBL; J00321; AAA98529.1; JOINED.  
DR EMBL; J00322; AAA98529.1; JOINED.  
DR EMBL; J00323; AAA98529.1; JOINED.  
DR EMBL; J00324; AAA98529.1; JOINED.  
DR EMBL; J00325; AAA98529.1; JOINED.  
DR EMBL; J00326; AAA98529.1; JOINED.  
DR EMBL; J00327; AAA98529.1; JOINED.  
DR EMBL; J00328; AAA98529.1; JOINED.  
DR EMBL; J00329; AAA98529.1; JOINED.  
DR EMBL; J00330; AAA98529.1; JOINED.  
DR EMBL; J00331; AAA98529.1; JOINED.  
DR EMBL; J00332; AAA98529.1; JOINED.  
DR EMBL; J00333; AAA98529.1; JOINED.  
DR EMBL; J00334; AAA98529.1; JOINED.  
DR EMBL; J00335; AAA98529.1; JOINED.  
DR EMBL; J00336; AAA98529.1; JOINED.  
DR EMBL; J00337; AAA98529.1; JOINED.  
DR EMBL; J00338; AAA98529.1; JOINED.  
DR EMBL; J00339; AAA98529.1; JOINED.  
DR EMBL; J00340; AAA98529.1; JOINED.  
DR EMBL; J00341; AAA98529.1; JOINED.  
DR EMBL; J00342; AAA98529.1; JOINED.  
DR EMBL; J00343; AAA98529.1; JOINED.  
DR EMBL; J00344; AAA98529.1; JOINED.  
DR EMBL; J00345; AAA98529.1; JOINED.  
DR EMBL; J00346; AAA98529.1; JOINED.  
DR EMBL; J00347; AAA98529.1; JOINED.  
DR EMBL; J00348; AAA98529.1; JOINED.  
DR EMBL; J00349; AAA98529.1; JOINED.  
DR EMBL; J00350; AAA98529.1; JOINED.  
DR EMBL; J00351; AAA98529.1; JOINED.  
DR EMBL; J00352; AAA98529.1; JOINED.  
DR EMBL; J00353; AAA98529.1; JOINED.  
DR EMBL; J00354; AAA98529.1; JOINED.  
DR EMBL; J00355; AAA98529.1; JOINED.  
DR EMBL; J00356; AAA98529.1; JOINED.  
DR EMBL; J00357; AAA98529.1; JOINED.  
DR EMBL; J00358; AAA98529.1; JOINED.  
DR EMBL; J00359; AAA98529.1; JOINED.  
DR EMBL; J00360; AAA98529.1; JOINED.  
DR EMBL; J00361; AAA98529.1; JOINED.  
DR EMBL; J00362; AAA98529.1; JOINED.  
DR EMBL; J00363; AAA98529.1; JOINED.  
DR EMBL; J00364; AAA98529.1; JOINED.  
DR EMBL; J00365; AAA98529.1; JOINED.  
DR EMBL; J00366; AAA98529.1; JOINED.  
DR EMBL; J00367; AAA98529.1; JOINED.  
DR EMBL; J00368; AAA98529.1; JOINED.  
DR EMBL; J00369; AAA98529.1; JOINED.  
DR EMBL; J00370; AAA98529.1; JOINED.  
DR EMBL; J00371; AAA98529.1; JOINED.  
DR EMBL; J00372; AAA98529.1; JOINED.  
DR EMBL; J00373; AAA98529.1; JOINED.  
DR EMBL; J00374; AAA98529.1; JOINED.  
DR EMBL; J00375; AAA98529.1; JOINED.  
DR EMBL; J00376; AAA98529.1; JOINED.  
DR EMBL; J00377; AAA98529.1; JOINED.  
DR EMBL; J00378; AAA98529.1; JOINED.  
DR EMBL; J00379; AAA98529.1; JOINED.  
DR EMBL; J00380; AAA98529.1; JOINED.  
DR EMBL; J00381; AAA98529.1; JOINED.  
DR EMBL; J00382; AAA98529.1; JOINED.  
DR EMBL; J00383; AAA98529.1; JOINED.  
DR EMBL; J00384; AAA98529.1; JOINED.  
DR EMBL; J00385; AAA98529.1; JOINED.  
DR EMBL; J00386; AAA98529.1; JOINED.  
DR EMBL; J00387; AAA98529.1; JOINED.  
DR EMBL; J00388; AAA98529.1; JOINED.  
DR EMBL; J00389; AAA98529.1; JOINED.  
DR EMBL; J00390; AAA98529.1; JOINED.  
DR EMBL; J00391; AAA98529.1; JOINED.  
DR EMBL; J00392; AAA98529.1; JOINED.  
DR EMBL; J00393; AAA98529.1; JOINED.  
DR EMBL; J00394; AAA98529.1; JOINED.  
DR EMBL; J00395; AAA98529.1; JOINED.  
DR EMBL; J00396; AAA98529.1; JOINED.  
DR EMBL; J00397; AAA98529.1; JOINED.  
DR EMBL; J00398; AAA98529.1; JOINED.  
DR EMBL; J00399; AAA98529.1; JOINED.  
DR EMBL; J00400; AAA98529.1; JOINED.  
DR EMBL; J00401; AAA98529.1; JOINED.  
DR EMBL; J00402; AAA98529.1; JOINED.  
DR EMBL; J00403; AAA98529.1; JOINED.  
DR EMBL; J00404; AAA98529.1; JOINED.  
DR EMBL; J00405; AAA98529.1; JOINED.  
DR EMBL; J00406; AAA98529.1; JOINED.  
DR EMBL; J00407; AAA98529.1; JOINED.  
DR EMBL; J00408; AAA98529.1; JOINED.  
DR EMBL; J00409; AAA98529.1; JOINED.  
DR EMBL; J00410; AAA98529.1; JOINED.  
DR EMBL; J00411; AAA98529.1; JOINED.  
DR EMBL; J00412; AAA98529.1; JOINED.  
DR EMBL; J00413; AAA98529.1; JOINED.  
DR EMBL; J00414; AAA98529.1; JOINED.  
DR EMBL; J00415; AAA98529.1; JOINED.  
DR EMBL; J00416; AAA98529.1; JOINED.  
DR EMBL; J00417; AAA98529.1; JOINED.  
DR EMBL; J00418; AAA98529.1; JOINED.  
DR EMBL; J00419; AAA98529.1; JOINED.  
DR EMBL; J00420; AAA98529.1; JOINED.  
DR EMBL; J00421; AAA98529.1; JOINED.  
DR EMBL; J00422; AAA98529.1; JOINED.  
DR EMBL; J00423; AAA98529.1; JOINED.  
DR EMBL; J00424; AAA98529.1; JOINED.  
DR EMBL; J00425; AAA98529.1; JOINED.  
DR EMBL; J00426; AAA98529.1; JOINED.  
DR EMBL; J00427; AAA98529.1; JOINED.  
DR EMBL; J00428; AAA98529.1; JOINED.  
DR EMBL; J00429; AAA98529.1; JOINED.  
DR EMBL; J00430; AAA98529.1; JOINED.  
DR EMBL; J00431; AAA98529.1; JOINED.  
DR EMBL; J00432; AAA98529.1; JOINED.  
DR EMBL; J00433; AAA98529.1; JOINED.  
DR EMBL; J00434; AAA98529.1; JOINED.  
DR EMBL; J00435; AAA98529.1; JOINED.  
DR EMBL; J00436; AAA98529.1; JOINED.  
DR EMBL; J00437; AAA98529.1; JOINED.  
DR EMBL; J00438; AAA98529.1; JOINED.  
DR EMBL; J00439; AAA98529.1; JOINED.  
DR EMBL; J00440; AAA98529.1; JOINED.  
DR EMBL; J00441; AAA98529.1; JOINED.  
DR EMBL; J00442; AAA98529.1; JOINED.  
DR EMBL; J00443; AAA98529.1; JOINED.  
DR EMBL; J00444; AAA98529.1; JOINED.  
DR EMBL; J00445; AAA98529.1; JOINED.  
DR EMBL; J00446; AAA98529.1; JOINED.  
DR EMBL; J00447; AAA98529.1; JOINED.  
DR EMBL; J00448; AAA98529.1; JOINED.  
DR EMBL; J00449; AAA98529.1; JOINED.  
DR EMBL; J00450; AAA98529.1; JOINED.  
DR EMBL; J00451; AAA98529.1; JOINED.  
DR EMBL; J00452; AAA98529.1; JOINED.  
DR EMBL; J00453; AAA98529.1; JOINED.  
DR EMBL; J00454; AAA98529.1; JOINED.  
DR EMBL; J00455; AAA98529.1; JOINED.  
DR EMBL; J00456; AAA98529.1; JOINED.  
DR EMBL; J00457; AAA98529.1; JOINED.  
DR EMBL; J00458; AAA98529.1; JOINED.  
DR EMBL; J00459; AAA98529.1; JOINED.  
DR EMBL; J00460; AAA98529.1; JOINED.  
DR EMBL; J00461; AAA98529.1; JOINED.  
DR EMBL; J00462; AAA98529.1; JOINED.  
DR EMBL; J00463; AAA98529.1; JOINED.  
DR EMBL; J00464; AAA98529.1; JOINED.  
DR EMBL; J00465; AAA98529.1; JOINED.  
DR EMBL; J00466; AAA98529.1; JOINED.  
DR EMBL; J00467; AAA98529.1; JOINED.  
DR EMBL; J00468; AAA98529.1; JOINED.  
DR EMBL; J00469; AAA98529.1; JOINED.  
DR EMBL; J00470; AAA98529.1; JOINED.  
DR EMBL; J00471; AAA98529.1; JOINED.  
DR EMBL; J00472; AAA98529.1; JOINED.  
DR EMBL; J00473; AAA98529.1; JOINED.  
DR EMBL; J00474; AAA98529.1; JOINED.  
DR EMBL; J00475; AAA98529.1; JOINED.  
DR EMBL; J00476; AAA98529.1; JOINED.  
DR EMBL; J00477; AAA98529.1; JOINED.  
DR EMBL; J00478; AAA98529.1; JOINED.  
DR EMBL; J00479; AAA98529.1; JOINED.  
DR EMBL; J00480; AAA98529.1; JOINED.  
DR EMBL; J00481; AAA98529.1; JOINED.  
DR EMBL; J00482; AAA98529.1; JOINED.  
DR EMBL; J00483; AAA98529.1; JOINED.  
DR EMBL; J00484; AAA98529.1; JOINED.  
DR EMBL; J00485; AAA98529.1; JOINED.  
DR EMBL; J00486; AAA98529.1; JOINED.  
DR EMBL; J00487; AAA98529.1; JOINED.  
DR EMBL; J00488; AAA98529.1; JOINED.  
DR EMBL; J00489; AAA98529.1; JOINED.  
DR EMBL; J00490; AAA98529.1; JOINED.  
DR EMBL; J00491; AAA98529.1; JOINED.  
DR EMBL; J00492; AAA98529.1; JOINED.  
DR EMBL; J00493; AAA98529.1; JOINED.  
DR EMBL; J00494; AAA98529.1; JOINED.  
DR EMBL; J00495; AAA98529.1; JOINED.  
DR EMBL; J00496; AAA98529.1; JOINED.  
DR EMBL; J00497; AAA98529.1; JOINED.  
DR EMBL; J00498; AAA98529.1; JOINED.  
DR EMBL; J00499; AAA98529.1; JOINED.  
DR EMBL; J00500; AAA98529.1; JOINED.  
DR EMBL; J00501; AAA98529.1; JOINED.  
DR EMBL; J00502; AAA98529.1; JOINED.  
DR EMBL; J00503; AAA98529.1; JOINED.  
DR EMBL; J00504; AAA98529.1; JOINED.  
DR EMBL; J00505; AAA98529.1; JOINED.  
DR EMBL; J00506; AAA98529.1; JOINED.  
DR EMBL; J00507; AAA98529.1; JOINED.  
DR EMBL; J00508; AAA98529.1; JOINED.  
DR EMBL; J00509; AAA98529.1; JOINED.  
DR EMBL; J00510; AAA98529.1; JOINED.  
DR EMBL; J00511; AAA98529.1; JOINED.  
DR EMBL; J00512; AAA98529.1; JOINED.  
DR EMBL; J00513; AAA98529.1; JOINED.  
DR EMBL; J00514; AAA98529.1; JOINED.  
DR EMBL; J00515; AAA98529.1; JOINED.  
DR EMBL; J00516; AAA98529.1; JOINED.  
DR EMBL; J00517; AAA98529.1; JOINED.  
DR EMBL; J00518; AAA98529.1; JOINED.  
DR EMBL; J00519; AAA98529.1; JOINED.  
DR EMBL; J00520; AAA98529.1; JOINED.  
DR EMBL; J00521; AAA98529.1; JOINED.  
DR EMBL; J00522; AAA98529.1; JOINED.  
DR EMBL; J00523; AAA98529.1; JOINED.  
DR EMBL; J00524; AAA98529.1; JOINED.  
DR EMBL; J00525; AAA98529.1; JOINED.  
DR EMBL; J00526; AAA98529.1; JOINED.  
DR EMBL; J00527; AAA98529.1; JOINED.  
DR EMBL; J00528; AAA98529.1; JOINED.  
DR EMBL; J00529; AAA98529.1; JOINED.  
DR EMBL; J00530; AAA98529.1; JOINED.  
DR EMBL; J00531; AAA98529.1; JOINED.  
DR EMBL; J00532; AAA98529.1; JOINED.  
DR EMBL; J00533; AAA98529.1; JOINED.  
DR EMBL; J00534; AAA98529.1; JOINED.  
DR EMBL; J00535; AAA98529.1; JOINED.  
DR EMBL; J00536; AAA98529.1; JOINED.  
DR EMBL; J00537; AAA98529.1; JOINED.  
DR EMBL; J00538; AAA98529.1; JOINED.  
DR EMBL; J00539; AAA98529.1; JOINED.  
DR EMBL; J00540; AAA98529.1; JOINED.  
DR EMBL; J00541; AAA98529.1; JOINED.  
DR EMBL; J00542; AAA98529.1; JOINED.  
DR EMBL; J00543; AAA98529.1; JOINED.  
DR EMBL; J00544; AAA98529.1; JOINED.  
DR EMBL; J00545; AAA98529.1; JOINED.  
DR EMBL; J00546; AAA98529.1; JOINED.  
DR EMBL; J00547; AAA98529.1; JOINED.  
DR EMBL; J00548; AAA98529.1; JOINED.  
DR EMBL; J00549; AAA98529.1; JOINED.  
DR EMBL; J00550; AAA98529.1; JOINED.  
DR EMBL; J00551; AAA98529.1; JOINED.  
DR EMBL; J00552; AAA98529.1; JOINED.  
DR EMBL; J00553; AAA98529.1; JOINED.  
DR EMBL; J00554; AAA98529.1; JOINED.  
DR EMBL; J00555; AAA98529.1; JOINED.  
DR EMBL; J00556; AAA98529.1; JOINED.  
DR EMBL; J00557; AAA98529.1; JOINED.  
DR EMBL; J00558; AAA98529.1; JOINED.  
DR EMBL; J00559; AAA98529.1; JOINED.  
DR EMBL; J00560; AAA98529.1; JOINED.  
DR EMBL; J00561; AAA98529.1; JOINED.  
DR EMBL; J00562; AAA98529.1; JOINED.  
DR EMBL; J00563; AAA98529.1; JOINED.  
DR EMBL; J00564; AAA98529.1; JOINED.  
DR EMBL; J00565; AAA98529.1; JOINED.  
DR EMBL; J00566; AAA98529.1; JOINED.  
DR EMBL; J00567; AAA98529.1; JOINED.  
DR EMBL; J00568; AAA98529.1; JOINED.  
DR EMBL; J00569; AAA98529.1; JOINED.  
DR EMBL; J00570; AAA98529.1; JOINED.  
DR EMBL; J00571; AAA98529.1; JOINED.  
DR EMBL; J00572; AAA98529.1; JOINED.  
DR EMBL; J00573; AAA98529.1; JOINED.  
DR EMBL; J00574; AAA98529.1; JOINED.  
DR EMBL; J00575; AAA98529.1; JOINED.  
DR EMBL; J00576; AAA98529.1; JOINED.  
DR EMBL; J00577; AAA98529.1; JOINED.  
DR EMBL; J00578; AAA98529.1; JOINED.  
DR EMBL; J00579; AAA98529.1; JOINED.  
DR EMBL; J00580; AAA98529.1; JOINED.  
DR EMBL; J00581; AAA98529.1; JOINED.  
DR EMBL; J00582; AAA98529.1; JOINED.  
DR EMBL; J00583; AAA98529.1; JOINED.  
DR EMBL; J00584; AAA98529.1; JOINED.  
DR EMBL; J00585; AAA98529.1; JOINED.  
DR EMBL; J00586; AAA98529.1; JOINED.  
DR EMBL; J00587; AAA98529.1; JOINED.  
DR EMBL; J00588; AAA98529.1; JOINED.  
DR EMBL; J00589; AAA98529.1; JOINED.  
DR EMBL; J00590; AAA98529.1; JOINED.  
DR EMBL; J00591; AAA98529.1; JOINED.  
DR EMBL; J00592; AAA98529.1; JOINED.  
DR EMBL; J00593; AAA98529.1; JOINED.  
DR EMBL; J00594; AAA98529.1; JOINED.  
DR EMBL; J00595; AAA98529.1; JOINED.  
DR EMBL; J00596; AAA98529.1; JOINED.  
DR EMBL; J00597; AAA98529.1; JOINED.  
DR EMBL; J00598; AAA98529.1; JOINED.  
DR EMBL; J00599; AAA98529.1; JOINED.  
DR EMBL; J00600; AAA98529.1; JOINED.  
DR EMBL; J00601; AAA98529.1; JOINED.  
DR EMBL; J00602; AAA98529.1; JOINED.  
DR EMBL; J00603; AAA98529.1; JOINED.  
DR EMBL; J00604; AAA98529.1; JOINED.  
DR EMBL; J00605; AAA98529.1; JOINED.  
DR EMBL; J00606; AAA98529.1; JOINED.  
DR EMBL; J00607; AAA98529.1; JOINED.  
DR EMBL; J00608; AAA98529.1; JOINED.  
DR EMBL; J00609; AAA98529.1; JOINED.  
DR EMBL; J00610; AAA98529.1; JOINED.  
DR EMBL; J00611; AAA98529.1; JOINED.  
DR EMBL; J00612; AAA98529.1; JOINED.  
DR EMBL; J00613; AAA98529.1; JOINED.  
DR EMBL; J00614; AAA98529.1; JOINED.  
DR EMBL; J00615; AAA98529.1; JOINED.  
DR EMBL; J00616; AAA98529.1; JOINED.  
DR EMBL; J00617; AAA98529.1; JOINED.  
DR EMBL; J00618; AAA98529.1; JOINED.  
DR EMBL; J00619; AAA98529.1; JOINED.  
DR EMBL; J00620; AAA98529.1; JOINED.  
DR EMBL; J00621; AAA98529.1; JOINED.  
DR EMBL; J00622; AAA98529.1; JOINED.  
DR EMBL; J00623; AAA98529.1; JOINED.  
DR EMBL; J00624; AAA98529.1; JOINED.  
DR EMBL; J00625; AAA98529.1; JOINED.  
DR EMBL; J00626; AAA98529.1; JOINED.  
DR EMBL; J00627; AAA98529.1; JOINED.  
DR EMBL; J00628; AAA98529.1; JOINED.  
DR EMBL; J00629; AAA98529.1; JOINED.  
DR EMBL; J00630; AAA98529.1; JOINED.  
DR EMBL; J00631; AAA98529.1; JOINED.  
DR EMBL; J00632; AAA98529.1; JOINED.  
DR EMBL; J00633; AAA98529.1; JOINED.  
DR EMBL; J00634; AAA98529.1; JOINED.  
DR EMBL; J00635; AAA98529.1; JOINED.  
DR EMBL; J00636; AAA98529.1; JOINED.  
DR EMBL; J00637; AAA98529.1; JOINED.  
DR EMBL; J00638; AAA98529.1; JOINED.  
DR EMBL; J00639; AAA98529.1; JOINED.  
DR EMBL; J00640; AAA98529.1; JOINED.  
DR EMBL; J00641; AAA98529.1; JOINED.  
DR EMBL; J00642; AAA98529.1; JOINED.  
DR EMBL; J00643; AAA98529.1; JOINED.  
DR EMBL; J00644; AAA98529.1; JOINED.  
DR EMBL; J00645; AAA98529.1; JOINED.  
DR EMBL; J00646; AAA98529.1; JOINED.  
DR EMBL; J00647; AAA98529.1; JOINED.  
DR EMBL; J00648; AAA98529.1; JOINED.  
DR EMBL; J00649; AAA98529.1; JOINED.  
DR EMBL; J00650; AAA98529.1; JOINED.  
DR EMBL; J00651; AAA98529.1; JOINED.  
DR EMBL; J00652; AAA98529.1; JOINED.  
DR EMBL; J00653; AAA98529.1; JOINED.  
DR EMBL; J00654; AAA98529.1; JOINED.  
DR EMBL; J00655; AAA98529.1; JOINED.  
DR EMBL; J00656; AAA98529.1; JOINED.  
DR EMBL; J00657; AAA98529.1; JOINED.  
DR EMBL; J00658; AAA98529.1; JOINED.  
DR EMBL; J00659; AAA98529.1; JOINED.  
DR EMBL; J00660; AAA98529.1; JOINED.  
DR EMBL; J00661; AAA98529.1; JOINED.  
DR EMBL; J00662; AAA98529.1; JOINED.  
DR EMBL; J00663; AAA98529.1; JOINED.  
DR EMBL; J00664; AAA98529.1; JOINED.  
DR EMBL; J00665; AAA98529.1; JOINED.  
DR EMBL; J00666; AAA98529.1; JOINED.  
DR EMBL; J00667; AAA98529.1; JOINED.  
DR EMBL; J00668; AAA98529.1; JOINED.  
DR EMBL; J00669; AAA98529.1; JOINED.  
DR EMBL; J00670; AAA98529.1; JOINED.  
DR EMBL; J00671; AAA98529.1; JOINED.  
DR EMBL; J00672; AAA98529.1; JOINED.  
DR EMBL; J00673; AAA98529.1; JOINED.  
DR EMBL; J00674; AAA98529.1; JOINED.  
DR EMBL; J00675; AAA98529.1; JOINED.  
DR EMBL; J00676; AAA98529.1; JOINED.  
DR EMBL; J00677; AAA98529.1; JOINED.  
DR EMBL; J00678; AAA98529.1; JOINED.  
DR EMBL; J00679; AAA98529.1; JOINED.  
DR EMBL; J00680; AAA98529.1; JOINED.  
DR EMBL; J00681; AAA98529.1; JOINED.  
DR EMBL; J00682; AAA98529.1; JOINED.  
DR EMBL; J00683; AAA98529.1; JOINED.  
DR EMBL; J00684; AAA98529.1; JOINED.  
DR EMBL; J00685; AAA98529.1; JOINED.  
DR EMBL; J00686; AAA98529.1; JOINED.  
DR EMBL; J00687; AAA98529.1; JOINED.  
DR EMBL; J00688; AAA98529.1; JOINED.  
DR EMBL; J00689; AAA98529.1; JOINED.  
DR EMBL; J00690; AAA98529.1; JOINED.  
DR EMBL; J00691; AAA98529.1; JOINED.  
DR EMBL; J00692; AAA98529.1; JOINED.  
DR EMBL; J00693; AAA98529.1; JOINED.  
DR EMBL; J00694; AAA98529.1; JOINED.  
DR EMBL; J00695; AAA98529.1; JOINED.  
DR EMBL; J00696; AAA98529

QY	86	KIYLCTPQEEFLEVDTSDDFWYVSVICKSNACKNHOREDPDKSSTFONLCKPISIHFG	145
Db	79	TIGISTPMDPTFVVPEDTSSNLNWPVSVCSSLACTNNHNRFNEDSDTSIOSTSEIYSITYG	138
QY	146	TGSMOGLIGDYDTVTVSNITVDIOQTGVLSTOEPGDVFTYAEPDGLIGMAVPSLASESIPV	205
Db	139	TGSMGTIGDYDTVQVGGIGISDFTMQIFGSLSETEPGEFLFYAAPFDGILGLAVPSISSGAPV	198
QY	206	FDNMANNRLHVAODLFSSYMDRNGQE-SMILTLGAIIDPSYTTGSLHWVPYTVQIWOPTYDS	264
Db	199	FDNIMNGLVSDLFSEVYLLASADDSGSSVYIFEGIDSSYTTGSLNWPVPYAVEGEWOITYDS	258
QY	265	VTSIGVVAAECGGCAIILIDPTGTRKGVGSSPDILIQOAIAGATONOXGFEFDICDNLSTWP	324
Db	259	ITMNEALACACGCAIYDITGTSLLTGFPSPLANIQSDIGASENDGDWVSCSAISLIP	318
QY	325	TWVEFINGKMPVLPPLSAVTSODQFCTSGFQSEN---HSQKWLGDVFIREFYSVFDRA	380
Db	319	DIVFINGVQVPPVPSAVIILQEGSCISGFGQMNLPTESGELMILGDVFIROQFYVFDRA	378
QY	381	NNLVGLA 387	
Db	379	NNQVGLA 385	
RESULT	10		
PEP4_MACEFU			
ID	PEP4_MACEFU	STANDARD:	PRT: 388 AA.
AC	P27678:		
DT	01-AUG-1992 (Rel. 23, Created)		
DT	01-AUG-1992 (Rel. 23, Last sequence update)		
DT	01-NOV-1995 (Rel. 32, Last annotation update)		
DE	Pepsin A-4 precursor (EC 3.4.23.1) (Pepsin I/II).		
GN	PGA.		
OS	Macaca fuscata fuscata (Japanese macaque).		
OC	Eukaryota; Metazoa; Chordata; Crenilata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;		
OC	Cercopitheciinae; Macaca.		
OX	NCHI_TaxID=9543;		
RN	[1]		
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 16-70.		
RC	TISSUE=Gastric mucosa;		
RX	MEDLINE=92037645; PubMed=1935977;		
RA	Kageyama T., Tanabe K., Koizumi O.;		
RT	"Development-dependent expression of isozymogens of monkey		
RT	pepsinogens and structural differences between them.";		
RL	Eur. J. Biochem. 202:205-215(1991).		
CC	-1- FUNCTION: SHOWS PARTICULARLY BROAD SPECIFICITY; ALTHOUGH BONDS		
CC	INVOLVING PHENYLALANINE AND LEUCINE ARE PREFERRED, MANY OTHERS ARE		
CC	ALSO CLEAVED TO SOME EXTENT.		
CC	-1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE PHE-, LEU-		
CC	-1- MISCELLANEOUS: THE EXPRESSION OF PEPsin GENES IS REGULATED BY		
CC	HORMONES AND RELATED SUBSTANCES.		
CC	-1- MISCELLANEOUS: EACH PEPsin GENES IS CONVERTED TO CORRESPONDING		
CC	PEPsin AT PH 2.0 IN PART AS A RESULT OF THE RELEASE OF A 47 AA		
CC	ACTIVATION SEGMENT AND IN PART AS A RESULT OF STEPWISE PROTEOLYTIC		
CC	CLEAVAGE VIA AN INTERMEDIATE FORM(S).		
CC	-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE		
CC	EUCHARIOTIC ASPARTYL PROTEASES FAMILY.		
CC	-----		
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation		
CC	at the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>		
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).		
CC	-----		
DR	EMBL: X59753: CAA42425.1; ..		
DR	PIR: S16065; S16065.		
DR	PIR: S19682; S19682.		
DR	HSSP: P00790: IPSN.		

Query	Local Similarity	Score	DB	Length
MEOPs; A01.001; -	52.1%;	1075.5;	DB 1;	388;
InterPro; IPR001969; Asp.protease.	56.6%;	Pred. No. 6.4e-75;		
InterPro; IPR001791; Laminin_G.	52;	Mismatches 99;	Indels 9;	Gaps 3
InterPro; IPR001461; Pepsin.				
Pfam; PF00026; asp. 1.				
PRINTS; PR00792; PEPsin.				
PROSITE; PS00141; ASP_PROTEASE; 2.				
Hydrolase; Aspartyl protease; Digestion; Stomach; Gastric juice;				
Zymogen; Multigene family; Signal; Glycoprotein.				
Signal	1	15		
PROPEP	16	38		
PROPEP	39	62		
CHAIN	63	388		
ACT_SITE	94	94		
ACT_SITE	277	277		
DISULFID	107	112		
DISULFID	268	272		
DISULFID	311	344		
CARBOHYD	88	88		
SEQUENCE	388 AA;	41955 MW;	A2923AB1F7FCDEB9	CRC64;

RT swine pepsinogen: involvement of the amino-terminal portion of the  
 RT activation peptide segment in restoration of the functional  
 RT protein.";  
 RL Gene 65:285-292(1988).  
 RN [2]  
 RP SEQUENCE OF 60-386.  
 RX MEDLINE=74299591; PubMed=4604255;  
 RA Moravsek L., Kostika V.;  
 RT "Complete amino acid sequence of hog pepsin.";  
 RL FEBS Lett. 43:207-211(1974).  
 RN [3]  
 RP SEQUENCE OF 16-134.  
 RX MEDLINE=74031413; PubMed=4584879;  
 RA Stepanov V.M., Baratova L.A., Pugacheva I.B., Belyanova L.P.,  
 RA Ravina L.P., Timokhina E.A.;  
 RT "N-terminal sequence of swine pepsinogen and pepsin. The site of  
 RT pepsinogen activation.";  
 RL Biochem. Biophys. Res. Commun. 54:1164-1170(1973).  
 RN [4]  
 RP SEQUENCE OF 16-56.  
 RX MEDLINE=69054241; PubMed=4881358;  
 RA Ong E.B., Perlmann G.E.;  
 RT "The amino-terminal sequence of porcine pepsinogen.";  
 RL J. Biol. Chem. 243:6104-6109(1968).  
 RN [5]  
 RP SEQUENCE OF 58-348.  
 RX MEDLINE=75211282; PubMed=1097438;  
 RA Sepulveda P., Marchinszyn J.P., Jr., Liu D., Tang J.;  
 RT "Primary structure of porcine pepsin. III. Amino acid sequence of a  
 RT cyanogen bromide fragment, CB2A, and the complete structure of  
 RT porcine pepsin.";  
 RL J. Biol. Chem. 250:5082-5088(1975).  
 RN [6]  
 RP PARTIAL SEQUENCE OF 1-26.  
 RX MEDLINE=6059312; PubMed=2415509;  
 RA Ichihara Y., Sogawa K., Takahashi K.;  
 RT "Isolation of human, swine, and rat prepepsinogens and calf  
 RT preprochymosin, and determination of the primary structures of their  
 RT NH2-terminal signal sequences.";  
 RL J. Biochem. 98:483-492(1985).  
 RN [7]  
 RP ACTIVE SITE.  
 RX MEDLINE=69283592; PubMed=4897201;  
 RA Bayliss R.S., Knowles J.R., Wybrandt G.B.;  
 RT "An aspartic acid residue at the active site of pepsin. The isolation  
 RT and sequence of the heptapeptide.";  
 RL Biochem. J. 113:377-386(1969).  
 RN [8]  
 RP X-RAY CRYSTALLOGRAPHY.  
 RX MEDLINE=78077917; PubMed=339692;  
 RA Andreeva N.S., Guschina A.E., Fedorov A.A., Shutzkever N.E.,  
 RA Volnova T.V.;  
 RT "X-ray crystallographic studies of pepsin.";  
 RL Adv. Exp. Med. Biol. 95:23-31(1977).  
 RN [9]  
 RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).  
 RX MEDLINE=90317821; PubMed=2115088;  
 RA Cooper J.B., Khan G., Taylor G., Tickle I.J., Blundell T.L.;  
 RT "X-ray analyses of aspartic proteinases. II. Three-dimensional  
 RT structure of the hexagonal crystal form of porcine pepsin at 2.3-A  
 RT resolution.";  
 RL J. Mol. Biol. 214:199-222(1990).  
 RN [10]  
 RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).  
 RX MEDLINE=91017500; PubMed=2217165;  
 RA Abad-Zapatero C., Rydel T.J., Erickson J.;  
 RT "Revised 2.3 A structure of porcine pepsin: evidence for a flexible  
 RT subdomain.";  
 RL Proteins 8:62-81(1990).  
 RN [11]  
 RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).  
 RX MEDLINE=91278095; PubMed=2056534;  
 RA Stielecki A.R., Fujinaga M., Read R.J., James M.N.G.;

RT "Refined structure of porcine pepsinogen at 1.8-A resolution.";  
 RL J. Mol. Biol. 219:671-692(1991).  
 RN [12]  
 RP X-RAY CRYSTALLOGRAPHY (1.65 ANGSTROMS).  
 RX MEDLINE=92279205; PubMed=1594574;  
 RA Hartsuck J.E., Koelsch G., Remington S.J.;  
 RT "The high-resolution crystal structure of porcine pepsinogen.";  
 RL Proteins 13:1-25(1992).  
 CC INVOLVING PHENYLALANINE AND LEUCINE ARE PREFERRED, MANY OTHERS ARE  
 CC ALSO CLEAVED TO SOME EXTENT.  
 CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE PHE- LEU-  
 CC -1- PTM: MINOR AMOUNTS OF THE ACTIVE ENZYME OCCUR WITH ALA-58 AT THE  
 CC AMINO END.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE  
 CC EURKAROTIC ASPARTYL PROTEASES FAMILY.  
 CC -1- DATABASE: NME-worthington enzyme manual;  
 CC WWW="http://www.worthington-biochem.com/manual/P/PM.html".  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M20920; AAA31095.1; -  
 DR PIR: A00982; PEPG.  
 DR PIR: B22434; B22434.  
 DR PIR: J70307; J70307.  
 DR PDB: 3PEP; 15-APR-90.  
 DR PDB: 4PEP; 15-JAN-91.  
 DR PDB: 5PEP; 15-JUL-90.  
 DR PDB: 1PSA; 31-JAN-94.  
 DR PDB: 2PSG; 15-OCT-92.  
 DR PDB: 3PSG; 15-JAN-93.  
 DR MEROPS: A01.001; -  
 DR InterPro: IPR001969; Asp-protease.  
 DR InterPro: IPR001791; Laminin-G.  
 DR InterPro: IPR001461; Pepsin.  
 DR Pfam: PF00026; asp. 1.  
 DR PRINTS: PR00792; PEPsin.  
 DR PROSITE: PS00141; ASP-PROTEASE; 2.  
 DR HydroLase: Aspartyl protease; Digestion; Stomach; Gastric juice;  
 KW Zymogen; Signal; Phosphorylation; 3d-structure.  
 KM SIGNAL.  
 FT SIGNAL 1 15  
 FT PROPEP 16 59  
 FT CHAIN 60 386  
 FT MOD\_RES 127 127  
 FT ACT\_SITE 91 91  
 FT ACT\_SITE 274 274  
 FT DISULFID 104 109  
 FT DISULFID 265 269  
 FT DISULFID 309 342  
 FT CONFLICT 34 34  
 FT CONFLICT 119 120  
 FT CONFLICT 128 128  
 FT CONFLICT 289 289  
 FT CONFLICT 314 314  
 FT CONFLICT 323 323  
 FT STRAND 61 64  
 FT STRAND 66 68  
 FT TURN 69 71  
 FT STRAND 72 79  
 FT TURN 80 83  
 FT STRAND 84 91  
 FT TURN 92 93  
 FT STRAND 97 99  
 FT STRAND 101 101  
 FT TURN 102 103  
 FT HELIX 107 110  
 FT TURN 111 111  
 N -> D (IN REF. 3 AND 4).  
 DS -> SD (IN REF. 2).  
 O -> E (IN REF. 2).  
 MISSING (IN REF. 2 AND 5).  
 S -> Q (IN REF. 5).  
 N -> D (IN REF. 5).









---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 31, 2002, 09:20:58 ; Search time 31.86 Seconds  
(without alignments)  
2117.640 Million cell updates/sec

Title: US-09-643-755b-2  
Perfect score: 2064  
Sequence: 1 MNFLKSPFYAFLCFGYFV.....REYYSVDRANLVGLAKAI 390

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues  
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL\_19:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1883	91.2	366	6	09NIP5	09NIP5 bubalus bub
2	1760.5	85.3	345	6	027951	027951 bos primige
3	1670	80.9	381	6	09GK11	09GK11 camelus dro
4	1601	77.6	380	6	028950	028950 sus scrofa
5	1423	68.9	379	11	09JUX1	09JUX1 ratus norv
6	1248	60.5	242	6	028075	028075 bos taurus
7	1105	53.5	386	6	09GMV6	09GMV6 canis famli
8	1099.5	53.3	385	13	09DEC4	09DEC4 rana catesb
9	1098	53.2	387	6	09GMV9	09GMV9 suncus muri
10	1090	52.8	387	6	09GMV8	09GMV8 sorex ungu
11	1087.5	52.7	385	6	029080	029080 sus scrofa
12	1087.5	52.7	386	6	09GMV7	09GMV7 rhinolophus
13	1084	52.5	390	6	09GK10	09GK10 camelus dro
14	1081	52.4	384	13	09DEC2	09DEC2 xenopus lae
15	1016.5	49.2	382	13	09PRG9	09PRG9 gallus gall
16	995	48.2	388	6	046524	046524 felis silve

17	970	47.0	388	6	046523	046523 equus zebra
18	962.5	46.6	378	13	09PUR9	09PUR9 pseudopleur
19	943	45.7	376	11	09PUR8	09PUR8 pseudopleur
20	922.5	44.7	387	11	09JUX2	09JUX2 ratus norv
21	918.5	44.5	387	11	09DIO6	09DIO6 mus musculu
22	918.5	44.5	387	11	09JKE6	09JKE6 mus musculu
23	899.5	43.6	389	6	09GMV5	09GMV5 suncus muri
24	894	43.3	383	13	09DEC3	09DEC3 xenopus lae
25	891.5	43.2	384	13	09I322	09I322 rana catesb
26	888	43.0	388	6	09GMV2	09GMV2 oryctolagus
27	880	42.6	392	11	09D7R7	09D7R7 mus musculu
28	878.5	42.6	389	6	09GMV4	09GMV4 sorex ungu
29	874.5	42.4	389	6	09GMV3	09GMV3 rhinolophus
30	863	41.8	387	13	09DDV5	09DDV5 salvelinus
31	859	41.6	383	13	09DE45	09DE45 salvelinus
32	850.5	41.2	389	13	09PMK1	09PMK1 gallus gall
33	849.5	41.2	389	13	09W643	09W643 gallus gall
34	846	41.0	370	6	09TTW1	09TTW1 bos taurus
35	846	41.0	375	6	046500	046500 bos taurus
36	845.5	41.0	387	6	046496	046496 bos taurus
37	843.5	40.9	406	13	P70068	P70068 pagothenia
38	827	40.1	396	11	09D7T2	09D7T2 mus musculu
39	822.5	39.8	389	6	09MYK3	09MYK3 sus scrofa
40	822	39.8	376	6	09TTY4	09TTY4 capra hircu
41	821.5	39.8	389	6	09MYK2	09MYK2 sus scrofa
42	821	39.8	397	6	028546	028546 ovis aries
43	812	39.3	386	6	09TTW8	09TTW8 capra hircu
44	801.5	38.8	399	13	093458	093458 podarcis si
45	792	38.4	402	13	057572	057572 chionodraco

## ALIGNMENTS

RESULT	ID	PRELIMINARY;	PRT;	366 AA.
09NIP5	09NIP5			
AC	09NIP5:			
DT	01-OCT-2000 (TREMBLrel. 15, Created)			
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)			
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)			
DE	PROCHYMOSIN.			
OS	Bubalus bubalis (Domestic water buffalo).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;			
OC	Bovidae; Bovinae; Bubalus.			
OX	NCBI_TaxID=89462;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Batish V.K., Mukhopadhyay U.K., Mohanty A.K., Grover S., Kuipers O.P.;			
RL	Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.			
CC	-1- SIMILARITY: BELONGS TO THE PEPTIDASE FAMILY A1; ALSO KNOWN AS THE			
CC	EUKARYOTIC ASPARTYL PROTEASES FAMILY.			
CC	EMBL: AF177290; AAF27315.1; -.			
DR	HSSP: P00794; 4CMS.			
DR	MEROPS: A01.006; -.			
DR	InterPro: IPR001969; Asp-Protease.			
DR	InterPro: IPR001461; Pepsin.			
DR	Pfam: PF00026; asp. 1.			
DR	PRINTS: PR00792; PEPsin.			
DR	PROSITE: PS00141; ASP_PROTEASE; 2.			
KW	Aspartyl protease; Hydrolase.			
SO	SEQUENCE 366 AA; 40498 MW; 279BED1258B5F15 CRC64;			

Query Match	91.2%;	Score 1883;	DB 6;	Length 366;
Best Local Similarity	97.8%;	Pred. No. 5,7e-150;		
Matches 357;	Conservative 3;	Mismatches 5;	Indels 0;	Gaps 0;
OY	26	AEIRIRPLIKKSLRKALKENGLEDFLOKQOYISISYSGFEVAVPLTNNYDSQYFG	85	
Db	2	AEIRIRPLIKKSLRKALKENGLEDFLOKQOYISISYSGFEVAVPLTNNYDSQYFG	61	

OY	86	KIYGTPPEETVLEFDGSDFWPVSITCKSNACKNRFDPKRSSTFQNLGRPLSHYG	145
Db	62	KIIYGTPPQEFTVLEFDGSSDFWVPSPITCKSNACKNNRDRPKRSSTFQNLGRPLSRIG	121
OY	146	TGSMOGLIGDYTVTSNIVDIQQTAVGLSTOBPGDVFYYAEFDGILGNAPYSLSAESSIPV	205
Db	122	TGSMOGLIGDYTAVTSNIVDIQQTAVGLSTOBPGDVFYYAEFDGILGNAPYSLSAESSIPV	181
OY	206	FDDMMNRHLVADLFPSVYMDBNGESMLTGADIDPSVTGSLLHNPVTVOOYNQFTVDSV	265
Db	182	FDDMMNRHLVADLFSVMDBNGESMLTLGADIDPSVTGSLHNPVTVOOYNQFTVDSI	241
OY	266	TISGVVACCEGCQAIIIDTGTSGTKLVGPSPSDLNTIOQAIGATONQYGEFDIDCDNLSTMP	329
Db	242	TISGVVACCEGCQAIIIDTGTSGTKLVGPSPSDLNTIOQAIGATONQYGEFDIDCDNLSTMP	301
OY	326	VVEFLNGKMYPLTFSATTSDQGCTGSGFSESNHSQKMLIGDVFIREYYSVFPDRANLVG	385
Db	302	VVEELNGMYPLTFSAYTSODQFGCTGSGFSENRSQQMILGDVFIREYYSVFPDRANLVG	361
OY	386	LAKAI 390	
Db	362	LAKAI 366	
RESULT	2		
ID	Q27951	PRELIMINARY;	PRT; 345 AA.
AC	Q27951:		
DT	01-NOV-1996 (TrEMBLrel_01, Created)		
DT	01-NOV-1996 (TrEMBLrel_01, Last sequence update)		
DT	01-DEC-2001 (TrEMBLrel_19, Last annotation update)		
DE	PROCHYMOSIN.		
OS	Bos primigenius (aurorochs).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;		
OC	Bovidae; Bos.		
OX	NCBI_TaxID=9909;		
RN	[1]		
RN	SEQUENCE FROM N.A.		
RP	Tan S.-Y., Zhang Y.-Y., Liu T.-Y., Liu N.-J., Yang K.-Y.;		
RA	"Nucleotide sequence of bovine prochymosin cDNA and the mechanism of		
RT	cDNA deletion."		
RL	Chin. J. Biotechnol. 5:328-332(1989).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RP	Xu L.;		
RA	Submitted (JAN-1995) to the EMBL/GenBank/DDBJ databases.		
RL	EMBL: U19786; AAA73476.1; -.		
DR	HSSP; P00794; 4CMS.		
DR	MEROPS; A01.006; -.		
DR	InterPro; IPR001969; Asp_protease.		
DR	InterPro; IPR001461; pepsin.		
DR	Pfam; PF00026; asp. 2.		
DR	PRINTS; PR00792; PEPSIN.		
DR	PROSITE; PS00141; ASP_PROTEASE; UNKNOWN_1.		
QC	SEQUENCE 345 AA; 38034 MW; 3960ECCE229661AFA CRC64;		

Query Match	85.3%	Score 1760.5;	DB 6;	Length 345;
Best Local Similarity	93.7%;	Pred. No. 1e-139;		
Matches 342;	Conservative 1;	Mismatches 1;	Indels 21;	Gaps 2.
QY	26	AEITPIPLYKGSUKRAKKEHGLEEDPLDQKQOYGISSKYSGCEVASVPLTNWLDLSQYRG	85	
DB	2	AEITRIPIPLYKGSUKRAK-KHLEEDPLDQKQOYGISSKYSGCEVASVPLTNWLDLSQYRG	60	
QY	86	KIYGTPEQEFVLVEDTSSDFWVPSIYCKSMACKNHQRFDRKKSFTPNLKKPLSIHHG	145	
DB	61	KIYGTPEQEFVLVEDTSS-----DKKSTFTPNLKKPLSIHHG	100	
QY	146	TGSMOGLIGYLVNYSNIVDIOQTWGLSTQERBGDYFYAAEPFGILIGMAPPSLASESTYV	205	

Query Match	Best Local Similarity	80.98;	Score 1670;	DB 6;	Length 381;
Matches	309;	Conservative 34;	Mismatches 28;	Indels 0;	Gaps
Db 101	TGSMGILGYDTVTVNSILTDIQGVIGISTGPPGGVFTYAEPDGILGMAFSLASEYSIPV	1655			
Qy 206	FDNMNRRLVAVQDFFSVYMDNRNGESMLTLGADIDPSTYTGSLHHVPTVQVYQVFTVDV	2655			
Db 161	FDNMNRRLVAVQDILLTSYMDNRNGESMLTLGADIDPSTYTGSLHHVPTVQVYQVFTVDV	2200			
Qy 266	TISGVVAVCEGGCAILDPTGTSKLVGPSISDLNTIOQAIGATONQYGEFDDICDNLSTMP	3252			
Db 221	TISGVVAVCEGGCAILDPTGTSKLVGPSISDLNTIOQAIGATONQYGEFDDICDNLSTMP	2800			
Qy 326	VFEINGKMYPLFPSAVYTSQDQGFCTSGFQSEHNSOKWILGADVIFIREYYSVFDNRANLV	3655			
Db 281	VFEINGKMYPLFPSAVYTSQDQGFCTSGFQSEHNSOKWILGADVIFIREYYSVFDNRANLV	3400			
Qy 386	LAKAI 390				
Db 341	LAKAI 345				
RESULT 3					
09GK11					
ID 09GK11	PRELIMINARY;	PRT;	381	AA.	
AC 09GK11;					
DT 01-MAR-2001 (TREMBLrel. 16, Created)					
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)					
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)					
DE CHYMOSIN PRECURSOR (EC 3.4.23.4).					
OS Camelus dromedarius (Dromedary) (Arabian camel).					
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
OC Mammalia; Eutheria; Cetartiodactyla; Tylopoda; Camelidae; Camelus.					
OX NCBI_TaxID=9838;					
LN [1]					
RP SEQUENCE FROM N.A.					
RC TISSUE=GASTRIC MUCOSA;					
RA Kappeler S.R., Farah Z., Puhan Z.;					
RT "Camel (Camelus dromedarius) Chymosin and Pepsin as Renneting Enzymes					
RT for Camel Milk.";					
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.					
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE					
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.					
DR EMBL: AJ131677; CAC19554.1; -.					
DR HSSP; P00794; 4CMS.					
DR MEROPS; A01.006; -.					
DR InterPro; IPR001969; Asp.protease.					
DR InterPro; IPR001461; Pepsin.					
DR Pfam; PF00026; asp. 1.					
DR PRINTS; PR00792; PEPsin.					
DR PROSITE; PS00141; ASP_PROTEASE; 1.					
DR Aspartyl protease; hydrolase; Signal.					
KM SIGNAL 1 58 POTENTIAL.					
FT CHAIN 59 381 CHYMOSIN.					
SQ SEQUENCE 381 AA; 42082 MW; 24BADB57B2E7FDD7 CRC64;					

[illegible]

Oy	260	FTVSVTISGVVAVCBECCQAIIIDTGTSKLVGPSSDLINQOALIGATOMONYGEFDDCDN	319
Db	251	FTVSVTINGVAACVGGCCAIIIDTGTSVLFGPSSDLIKIOMAGIATENRYGEEFDNCGN	310
Oy	320	LSVPTVFVEELNGMYLPLTPSAVYSODOGCTSGFQSENHOKMIIGDVIRREYVSFDR	379
Db	311	LRSMPVTFVEELNGDYDLPSPASVSKDPGCTSGFOGDNNSELMILGDVIRREYVSFDR	370
Oy	380	ANNLVGLAKAI 390	
Db	371	ANNRVGLAKAI 381	

RESULT	4			
028950				
ID	Q28950	PRELIMINARY;	PTT;	380 AA.
AC	Q28950;			
DT	01-NOV-1996 (TREMBLrel. 01, Created)			
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)			
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)			
DE	PREPROCHYMOSIN PRECURSOR (FRAGMENT).			
OS	Sus scrofa (Pig).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.			
OX	NCBI_TaxID=9823;			
PN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=GASTRIC;			
RA	Follmann B., Jensen A.L., Loenblad P., Smidt E., Axelsen N.H.;			
RT	"A developmental analysis of the production of chymosin and pepsin in			
RT	pigs.";			
RL	Comp. Biochem. Physiol. 68B:9-13(1981).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=GASTRIC;			
RX	MEDLINE=96252892; PubMed=8673731;			
RA	Houen G., Madsen M.T., Harlow K.W., Loenblad P., Follmann B.;			
RT	"The primary structure and enzymic properties of porcine prochymosin			
RT	and chymosin.";			
RL	Int. J. Biochem. Cell Biol. 28:667-675(1996).			
EMBL	U14406; AAB08492.1; -.			
DR	HSSP; P00794; 4CMS.			
DR	MEROPE; A01_006; -.			
DR	InterPro; IPR001969; Asp_protease.			
DR	InterPro; IPR001461; Pepsin.			
DR	Pfam; PF000026; asp. 1.			
DR	PRINTS; PR00792; PEPStn.			
DR	PROSITE; PS00141; ASP_PROTEASE; UNKNOWN_2.			
KW	Signal.			
FT	NON_TER	1	1	
FT	SIGNAL	<1	15	POTENTIAL.
FT	PROPEP	16	57	ACTIVATION PEPTIDE (POTENTIAL).
FT	CHAIN	58	380	CHYMOSIN.
QO	SEQUENCE	380 AA;	41771 MW;	ECA81B78239D2102 CRC64;

Query Match	77.6%;	Score 1601;	DB 6;	Length 380;
Best Local Similarity	80.1%;	Pred. No. 2.9e-126;		
Matches 297;	Conservative 35;	Mismatches 39;	Indels 0;	Gaps 0;

[illegible]

Qy	200	EYLPVDDNNMNNHIVLAQDLFAYVMRDNCOEMLTLGLIDPSYTGSLAHWPPLVWQIQYQ	255
	190	EYLPVDDNNMNNHIVLAQDLFAYVMRDNCOEMLTLGLIDPSYTGSLAHWPPLVWQIQYQ	249
Qy	260	FTVDSVITSGVNVVACCEGCGCALIIDPTGTSKLYVPSDDIINIOALIGATONQYGEFDIDCDN	319
	250	FTVDSVITSGVNVVACCEGCGCALIIDPTGTSKLYVPSDDIINIOALIGATONQYGEFDIDCDN	309
Db	320	LSIYPIVVEEINGKMYPLPLPSAVTSODOGFCTSGFQSENHSSOKWILGDVPLIREYYSVFR	379
Qy	310	LSIYPIVVEEINGKMYPLPLPSAVTSODOGFCTSGFQSENHSSOKWILGDVPLIREYYSVFR	369
Db	380	ANNLVGLAKAI 390	
Qy	370	ANNLVGLAKAI 380	

RESULT	5			
09JUX1				
ID	09JUX1	PRELIMINARY:	PRT:	379 AA.
AC	09JUX1;			
DT	01-Oct-2000 (TrEMBLrel. 15, Created)			
DT	01-Oct-2000 (TrEMBLrel. 15, Last sequence update)			
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)			
DE	PROCHYMOSIN PRECURSOR (EC 3.4.23.4).			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX	NCBI_TaxId=10116;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=SD;			
RX	MEDLINE=00139732; PubMed=10673373;			
RA	Kageyama T., Ichinose M., Tsukada Kato S., Omata M., Narita Y.,			
RA	Moriyama A., Yonezawa S.;			
RT	"Molecular cloning of neonate/infant-specific pepsinogens from rat			
RT	stomach mucosa and their expressional change during development.";			
RL	Biochem. Biophys. Res. Commun. 267:806-812(2000).			
DR	EMBL; AJ251688; CAB75983.1; -			
DR	HSSP; P00794; 4CMS.			
DR	MEROPS; A01.006; -;			
DR	InterPro; IPR001969; Asp_protease.			
DR	InterPro; IPR001461; Pepsin.			
DR	Pfam; PF000026; asp. 1.			
DR	PRINTS; PR00792; PEPsin.			
DR	PROSITE; PS00141; Asp_PROTEASE; UNKNOWN_2.			
KW	Signal; Hydrolase.			
FT	SIGNAL			
FT	CHAIN	1	16	POTENTIAL.
FT	CHAIN	17	379	PROCHYMOSIN.
SQ	SEQUENCE	379 AA; 42434 MW; 439650CCCT59A5ADC CRC64;		

Query Match	68.9%	Score 1423	DB 11	Length 379
Best Local Similarity	71.4%	Pred. No. 2.6e-111		
Matches 265	Conservative 42	Mismatches 62	Indels 2	Gaps 1

QY	20	VAAHAAETTRIPLYKSGSLKALKEHGLLEDPELOKOQOYGSXKSGCFEYASVPLNTYL	79
Db	11	LAAGSIVVTRIPLEHNGKSLRNTLLEQGLLEDPELRKHRYEFSEKNSNMGWASEPLNTYL	70
QY	80	DSQYFGKIVLGTIPPOEFYVLTPEFTGSSDEFWPESYCKSNACKNHOFRPDRKSFPTONLGRP	139
Db	71	DSEYFGIVVGPFPQEFKAYVPTDQSSSELWVPSYVCSKVCNHNRRPDRSKSFPTONLSK	130
QY	140	LSHYGTGSMOGLIGLDIVYTVYSNIYDIDQOYVGLSTQEPEDVFTYAEFDGILGMAYPSLAS	199
Db	131	LFQOYGTGSEVEGLAYDVFTVSDIYVAPQVGLSTEEERGLFTYSPDGILGLAYPTFAS	190
QY	200	EYSIPEFDNMNMNRHLVLAODLFVYMDRNGQSMTLIGIDPSYTGSLHWPVTVQOYQQ	259
Db	191	KYSPLEFDNMNMNRHLVLAODLPFVYMSRNDQSMTLIGAIDQSYTGSLHWPVTVQOYQW	250
QY	260	FTVDSVLTISGVVACEGGQALIDLTGTSLKLVGPSSDILNIDQALIGATONQYGEFDDICDN	319





OX NCBI\_TaxID=8400;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Ikuzawa M., Inokuchi T., Kobayashi K., Yasumasa S.;  
RT "Molecular Cloning of peptinogens in Adult Xenopus laevis and Bullfrogs  
Rana catesbeiana."  
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE  
EUKARYOTIC ASPARTYL PROTEASES FAMILY.  
DR EMBL: AB045376; BAB20092.1; -  
DR HSSP: P00794; 4CSN.  
DR MEROPS: A01.001; -  
DR InterPro: IPR001969; Asp\_protease.  
DR pfam: PF00026; asp. 1.  
DR PRINTS: PR00792; PEPsin.  
DR PROSITE: PS00141; ASP\_PROTEASE; 1.  
KW Aspartyl protease; Hydrolase.  
SQ SEQUENCE 385 AA; 41702 MW; 5DC8914FC9CEC603 CRC64;

Query Match 53.3%; Score 1099.5; DB 13; Length 385;  
Best Local Similarity 53.9%; Pred. No. 3.9e-84;  
Matches 213; Conservative 47; Mismatches 112; Indels 9; Gaps 4;

QY 13 LCFQGYEAVVTHAAETIRPIPLYKSKLRALKKEHGLEDFLOKQOYGISKY-SGFGEVA 71  
DB 5 LIFG---LVLAIEGCVAVSLRKGSLSARLNRLGLDGLKHHYNPAIKTFPSLAQNS 61  
QY 72 SVPLTNLDISOYFGKITYLGPPOEFTVLEDTGSSDFWVPSIYCKSNACKNHQRPDKRS 131  
DB 62 GEPLQNYMDIEFGTISIGTPOSTVFIEDTGSSNLWVPSYCSSPACNNHMFDPQSS 121  
QY 132 TFOQLKRLPSLHYTGSGNOGLGYDTVYSNIVDIQOTVGLSTQEPDVFYAEFDGILG 191  
DB 122 TFOQNTNPVSIOYTGSGSGELGYDTVOVGNITNQITFGLSQSPSGFLYXSPDGILG 181  
QY 192 MAYSLASEYSIPVFNMMNHLVLAQDLFSVYMDRNGOE-SMLTGAIDPSYYSGLHWY 250  
DB 182 LAFSLASLSQATPFVFNMMNGLIPQDLFSYLLSSQSGSVLEFGVDTSYTTNLMNV 241  
QY 251 PTVVQOITWQFTVDSVTSISGVVAVCEGCOAILDTGTSLKLVGSSDILNIOAIGATONQY 310  
DB 242 PLTAETVQITVDSISIGVYIACSGSCSAIVDTGTSLAGSTPIANIQYIYGANQNS 301  
QY 311 GEPLIDCDNLSYMPVVEINGKMPILTPSAVTSODGFCISGPOSEN----HSQKALIG 366  
DB 302 GOYVINCNNISMPVYFTINGVYPLPASAVYRQSQSCSTSGFQAMNLPRTSSGDLWILG 361  
QY 367 DVFTREYYSVFDNRANNLVGLA 387  
DB 362 DVFTREYYSVFDNRANNYAMA 382

RESULT 9  
O9GMV9 PRELIMINARY; PRT; 387 AA.  
AC O9GMV9  
DT 01-MAR-2001 (TREMblrel. 16, Created)  
DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)  
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
DE PEPsinogen A.  
GN PGNA.  
OS Sorex unguiculatus (House shrew) (Musk shrew).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Insectivora; Soricidae; Suncus.  
OX NCBI\_TaxID=9378;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Narita Y., Oda S., Takenaka O., Kageyama T.;  
RT "Phylogenetic position of Insectivora inferred from the cdna sequences  
of peptinogen A and C."  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.

CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE  
EUKARYOTIC ASPARTYL PROTEASES FAMILY.  
DR EMBL: AB047243; BAB11749.1; -  
DR HSSP: P00790; IFSN.  
DR InterPro: IPR001969; Asp\_protease.  
DR pfam: PF00026; asp. 1.  
DR PRINTS: PR00792; PEPsin.  
DR PROSITE: PS00141; ASP\_PROTEASE; 1.  
KW Aspartyl protease; Hydrolase.  
SQ SEQUENCE 387 AA; 41576 MW; 7F9F818DD541CE CRC64;

Query Match 53.2%; Score 1098; DB 6; Length 387;  
Best Local Similarity 57.1%; Pred. No. 5.3e-84;  
Matches 209; Conservative 56; Mismatches 93; Indels 8; Gaps 3;

QY 30 RIPLVYKKSILRKALKEHGLEDFLOKQOYGISKY---SGFGEVASVPLTNLDISOYFGK 86  
DB 19 KVIPLVKKSLRQNLLENGLDKDFLAKHNVPASKTFPTEATLEADQPLVNTMDMEYFCT 78  
QY 87 IYLGTPPOEFTVLEDTGSSDFWVPSIYCKSNACKNHQRPDKRSSTFQNLKPLSIHYGT 146  
DB 79 IGITPPPOEFTVIFDTGSSNLWVPSYCSSPACNNHNFDPQSSSTFQSTQTLSTAYCT 138  
QY 147 GSMOGLGYDTVYSNIVDIQOTVGLSTQEPDVFYAEFDGILGMAVSLASEYSIPV 206  
DB 139 GSMTGVLGIDTVQVAGIATNQITFGLSQEPGSPFLYXSPDGILGLAYINLASSGATPVF 198  
QY 207 DNMNMRHLVAQDLFSVYMDRNGOE-SMLTGAIDPSYYSGLHWYPTVQOITWQFTVDSV 265  
DB 199 DNMNNGVLVSQDLFSYLLSSNDGSSVYIFGIDSSYTYGNLWVPLSEGTWQITVDSI 258  
QY 266 TISGVVAVCEGCOAILDTGTSLKLVGSSDILNIOAIGATONQYGEPLIDCDNLSYMP 325  
DB 259 TFMGCAIACSGSCSAIVDTGTSLLGPNNAIANIQKISASONANQAMVAVSCSIOSLDP 318  
QY 326 VYFEINGKMPILTPSAVTSODGFCISGPOSEN---HSQKMLGDFVFTREYYSVFDNRAN 381  
DB 319 IYFTINGIYPLPASAVYIILNQODCTSGFQGMDFPSPGELWILGDFVTRQYFAVFDKSN 378  
QY 382 NLVGLA 387  
DB 379 NRVLGLA 384

RESULT 10  
O9GMV8 PRELIMINARY; PRT; 387 AA.  
AC O9GMV8  
DT 01-MAR-2001 (TREMblrel. 16, Created)  
DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)  
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
DE PEPsinogen A.  
GN PGNA.  
OS Sorex unguiculatus (Long-clawed shrew).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Insectivora; Soricidae; Sorex.  
OX NCBI\_TaxID=62275;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Narita Y., Oda S., Takenaka O., Kageyama T.;  
RT "Phylogenetic position of Insectivora inferred from the cdna sequences  
of peptinogen A and C."  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE  
EUKARYOTIC ASPARTYL PROTEASES FAMILY.  
DR EMBL: AB047244; BAB11750.1; -  
DR HSSP: P00790; IFSN.  
DR MEROPS: A01.001; -  
DR InterPro: IPR001969; Asp\_protease.  
DR pfam: PF00026; asp. 1.





```

QY 207 DNMMNRHLVAODLFSSVYMDRNGOE-SMLTLGAIDPSYTYGSLHWVPVTVQOYMOFTVDSV 265
    |||:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 196 DNMMNSQHLVAODLFSSVYLSKDGFTSGSVFLFGVDNYSYSSLMNVPPLTAETVWQITLDSV 255
QY 266 TISGVVACGEGCOAIIIDTGSTSKLVGPSDDLNTQOAIGATONOGYGEFIDCDNLSTYPT 325
    |||:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 256 SINGOVYACSSQSOCAIVDTGTSMTGSPFIANTQNYIGASODSNGQYVINCNNISNMP 315
QY 326 VVEINGKMYPLTPSAVYTSODGFCFTSGFOS----ENHSQKWLGDVFIREFYVSVPDRAN 381
    |||:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 316 IVTVINGVQPLPSASVYVRQNOGCGSSGFQAMNLPINSGLMWILGDVFIQYFTVFPDRAN 375
QY 382 NNVLGDA 387
    |||:::
Db 376 NYVAIA 381

RESULT 15
Q9PRG9 PRELIMINARY; PRT; 382 AA.
ID Q9PRG9:
AC Q9PRG9:
DT 01-MAY-2000 (T-EMBLrel. 13, Created)
DT 01-MAY-2000 (T-EMBLrel. 13, last sequence update)
DT 01-DEC-2001 (T-EMBLrel. 19, last annotation update)
DE PEPSINOGEN A.
GN CPGA.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
OX 11]
RP SQUENCE FROM N.A.
RA Sakamoto N., Yasugi S.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN 12]
RP SQUENCE FROM N.A.
RX MEDLINE=98440813; PubMed=9753645;
RA Sakamoto N., Saiga H., Yasugi S.;
RT "Analysis of temporal expression pattern and cis-regulatory sequences
    of chicken pepsinogen A and C."
RC Biochem. Biophys. Res. Commun. 250:420-424(1998).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
    EUKARYOTIC ASPARTYL PROTEASES FAMILY.
DR EMBL: AB025283; BAA77268.1; -.
DR EMBL: AB025281; BAA76891.1; -.
DR HSP; P00794; 4CMS.
DR MEROPS; A01.0PW; -.
DR InterPro; IPR001969; Asp_protease.
DR InterPro; IPR001461; Pepsin.
DR Pfam; PF00026; asp. 1.
DR PRINTS; PR00792; PEPSTN.
DR PROSITE; PS00141; ASP_PROTEASE; 2.
KW Aspartyl protease; Hydrolase.
SQ SEQUENCE 382 AA; 42056 MW; C3C7EB8CA4DD3E28 CRC64;

```

Query Match 49.2%; Score 1016.5; DB 13; Length 382;  
 Best Local Similarity 54.6%; Pred. No. 3.6e-77;  
 Matches 201; Conservative 57; Mismatches 97; Indels 13; Gaps 7;

```

QY 28 ITRPIPKKSLRKALKEHGLLEDFLOKQOYGISKXSGF--GEVASVPLFNLYDSQYFG 85
    |||:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 17 IHRVPLKKGSKLRQKLDHGLLEDFLKKHPYNPAKXTHPVLTATESEPMNTNYMDASTYG 76
QY 86 KVLGTPPEFTVLFDTGSSDFWVPSIYCKSMACKNHQREDPRKSTFFQNLGKPLSIHYG 145
    |||:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 77 TISIGTFQODFTVFDGSSNLAWPSIYCKSSACSNHKKRFPSPKSYVSTNETVYIAYG 136
QY 146 TGSNOGLIGDYTVASNIVIDQ-QTVGLSTDEPQDVFTYAFEDGILGMAYPSLASYSIP 204
    |||:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 137 TGSMSGILGDTYAVVSSI-DVQNOIFGLSETPEGSFFYCNFDSILGLAFPSISSGATP 195

```

```

QY 205 VFDNMNRHLVAODLFSSVYMDRNGOE-SMLTLGAIDPSYTYGSLHWVPVTVQOYMOFTVDSV 263
    |||:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 196 VFDNMNSQHLVAODLFSSVYLSKDGFTSGSVFLFGVDNYSYSSLMNVPPLTAETVWQITLDSV 255
QY 264 SVTISGVVACGEGCOAIIIDTGSTSKLVGPSDDLNTQOAIGATONOGYGEFIDCDNLSTYPT 323
    |||:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 256 RVTVGNKRYVACFFTCOAIVDTGTSMTGSPFIANTQNYIGASODSNGQYVINCNNISNMP 311
QY 324 PTVVEINGKMYPLTPSAVYTSODGFCFTSGFOS----ENHSQKWLGDVFIREFYVSVPDRAN 379
    |||:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 312 PDVTFHNGHAFPLPASAVYVINEGSCMLGFENNGTTELGQWILGDVFIREFYVIFDR 371
QY 380 ANNVLGDA 387
    |||:::
Db 372 ANNKVGLS 379

```

Search completed: July 31, 2002, 09:24:51  
 Job time: 233 sec



**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 18:18:54 ; Search time 9237.8 Seconds  
(without alignments)  
8963.856 Million cell updates/sec

Title: US-09-643-755B-3  
3957  
Perfect score: 1 ctcgagcaatcattgfract.....acctaccactaagttacc 3957  
Sequence:

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues  
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: gb\_ba:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_inv:\*  
32: em\_htg\_other:\*  
33: em\_htgo\_inv:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
------------	-------	-------------	--------	-------	-------------

1	3957	100.0	3957	6	AX088021	AX088021 Sequence
2	1551	39.2	3502	8	PHYCSP	J01263 Phaeosolus v
3	1353	34.2	1415	6	AX343913	AX343913 Sequence
4	1173	29.6	1173	6	AX088019	AX088019 Sequence
5	1149.8	29.1	2970	6	AX252300	AX252300 Sequence
6	900.4	22.8	1305	4	BOVCHYMOB	J00003 bovine chym
7	897.4	22.7	1240	6	AR002347	AR002347 Sequence
8	897.2	22.7	1460	6	E000042	E00042 DNA coding
9	896.2	22.6	1291	6	A15836	A15836 chymosin ge
10	895.8	22.6	2733	6	AR073077	AR073077 Sequence
11	895.6	22.6	1175	6	I04058	I04058 Sequence 5
12	895.5	22.6	2726	6	I08097	I08097 Sequence 5
13	894	22.6	1275	6	BOVCHYMOA	J00002 bovine chym
14	894	22.6	1290	6	E000075	E00075 CDNA encod
15	894	22.6	1291	6	E000075	E00075 CDNA encod
16	893	22.6	1289	6	E00144	A15533 preprochymo
17	892.4	22.6	1311	6	E00108	E00144 CDNA encod
18	892.4	22.6	1460	6	E00295	E00295 CDNA encod
19	890.8	22.5	1269	6	BOVCHYMOA	J00004 bovine chym
20	886.8	22.4	1098	6	E00079	E00079 CDNA sequen
21	886.8	22.4	1098	6	E02341	E02341 CDNA sequen
22	875.4	22.1	2382	6	I08098	I08098 Sequence 1
23	870.8	22.0	1101	4	AF177290	AF177290 Bubalus a
24	839.6	21.2	1292	4	OAPPCHY	X53037 Lamb mRNA f
25	776	19.6	1115	12	SYNPROCA	M22593 Synthetic b
26	753.6	19.0	1094	4	BPJ19786	U19786 Bos primige
27	749	18.9	1172	4	CDRI31677	AJ131677 Camelus d
28	734.4	18.6	1130	4	SS014406	U14406 Sus scrofa
29	718	18.1	1262	9	AB038386	AB038386 Callithr1
30	627.2	15.9	969	6	A02000	A02000 Artificial
31	627.2	15.7	969	6	A06446	A06446 Artificial
32	619.6	15.7	1146	12	SYNPROCA	M31016 Synthetic b
33	605	15.3	1259	10	RNO251688	AJ751688 Rattus no
34	534	13.5	1140	6	AX319837	AX319837 Sequence
35	383.4	9.7	1403	4	AB047246	AB047246 Canis fam
36	382.4	9.7	637	6	E05472	E05472 DNA sequenc
37	378.6	9.6	1161	4	AB047245	AB047245 Rhinolph
38	373.8	9.4	1351	9	AB038384	AB038384 Callithr1
39	370.8	9.4	1381	4	AB047244	AB047244 Sorex ung
40	364.6	9.2	1363	4	PTGPEPA	J04601 Pig pepsi
41	363.8	9.2	1362	4	AB047243	AB047243 Suncus mu
42	363.4	9.2	1221	4	CDRI31678	AJ131678 Camelus d
43	363.4	9.2	1276	5	CHKPSN	D00215 Gallus gall
44	361.6	9.1	1266	4	RABPEPT23	M59235 Rabbit peps
45	357.4	9.0	1353	9	MPPEPA23	X59755 M.fusca m

## ALIGNMENTS

RESULT 1	AX088021	3957 bp	DNA	linear	PAT 17-MAR-2001
LOCUS	AX088021	Sequence 3 from Patent WO0114571.			
DEFINITION	AX088021				
ACCESSION	AX088021.1	GI:13396949			
VERSION					
KEYWORDS					
SOURCE					
ORGANISM					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
FEATURES					
source					

1..3957  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
1554..2726  
/note="Figure 2"  
/codon\_start=1  
/transl\_table=11

/protein\_id="CAC34681.1"  
/db\_xref="GI:13396950"  
/translation="MNFLEKSPFYAFLFCGOYEVAVTHAEIRIPILYKSKSLKALK  
EHGLEDFILKQOQGISKXSGEVAFLNLNDQSGKILGLPDPFVLEPT  
GSSDFWPSLYCKSNACKNHOFRPSKSRFONLNGSLYHSGMOGLGYPTVY  
SNVIDQOQVIGISNOEPGDVETVAEPDGIIGMAVPSIASEYIPVPMNNRHVAD  
LFESYMDRNOESMLTIGALIDPSYTGSLRHVPYVOQYNOFPYDSTISGVYACGS  
GCOAILDPTGSKLVGPFSSDILNIOAIGATONQGERDIDCDNLSTMPYVFEIENK  
YPLPSAVYTSODQFCFSGFESENHSQKILIGDVFIREYYSVFEDRANNLGLAKAI"  
BASE COUNT 1263 a 790 c 609 g 1295 t  
ORIGIN

Query Match 100.0%; Score 3957; DB 6; Length 3957;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 3957; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctgcaagaatcatctgtactccagatcatatagtaagaagtttgctctcctgcgcg 60  
DB 1 CTGCGAAGATTCATTGTACTCCAGTATCATATAGTAAGATTTTGGCTCTCGCCGG 60  
QY 61 tggattttaccctattaaagggtttccacctaaaattctgtaatacttcact 120  
DB 61 TGGTTTTTACCTCTATTTAAAGGGGTTTCCACCTMAAAATTCGTAATCTCTCACT 120  
QY 121 ttactgttacttaatttcaataatcttggttgaattatcaagcttcgcgacaaga 180  
DB 121 TTACTTTGTTACTTTAATTTCTCATATCTTGGTTGAATTTATCAGCTTCGCGACAGA 180  
QY 181 tatccctacaattatatttgttaacaatttcaaacccgataaatttatgaagtc 240  
DB 181 TATCCCTACAAATTTATTTATTTGTTAAACATTTTCAAAACGCGATMAAATTTATGAAGTC 240  
QY 241 ccgtctatcttaagttagtcaacatttcaatttgaataataataatttacttaatt 300  
DB 241 CCGCTATCTTTAATGTAGTCTTACATTTTCAATTTGAATATATATTTACTTTAATTT 300  
QY 301 agcgtgtgtagaagaacaaattatcttacttcttccatabaaagtttaataata 360  
DB 301 ACGGTTGTGAGAAACATMAAGATTTATCTTCTCATATMAATGTTTATATA 360  
QY 361 caataaacaatttctacttaagaagaattcccaattatattttaaataat 420  
DB 361 CAATATMAAACAAATCTTAACTTAAAGAGATTTCCATTTATATTTTAAATAATAT 420  
QY 421 ttataaatatttcaaacacgtaaatctcaataaagaattggttcaaaagtaata 480  
DB 421 TTATCAAAATTTTCAACACGTAATCTCATTAATATATAGTTGTTCAAAAGTATATA 480  
QY 481 aatttaactcaaattttttacttgcactgactcttaagaacacccagtgacaaact 540  
DB 481 AATTTAACTCAATATTTTATTCGACTGATCTTAAAGCAACCCAGTGACACAAGCT 540  
QY 541 agccatttttcttgaataaaaaaatacaattatcatgtatattttttatagaaga 600  
DB 541 ACCCATTTTCTTGGATTAATAAAAAATCCAAATATCATGTATTTTATATACATGA 600  
QY 601 aaattcaccaacaacatcttggtgatttctgaagcaagtcagctttagcaaaattct 660  
DB 601 AATTTTCAACCAACATCTTCTGTATTTCTGAAGCAAGTCATTTATGCAAAATTTCT 660  
QY 661 ataattcccatctgacactacggaagtaaacgtgaagctcgctttagcatgagacacat 720  
DB 661 ATAAATCCCATTTGACACTACGGAAGTAAGTATGCTTTTATCATATGCGAGACACAT 720  
QY 721 cctctaaagtaattttaaataagtagttactatatacaagattcatatcaaatattcaa 780  
DB 721 CTTCTAAAGTAATTTAATATATAGTTACTATATTCAAAGATTTATATCAAAATFACCAA 780  
QY 781 tattacttaaaaaataatagatataataataattacttttaatttaagttt 840  
DB 781 TATTTACTCTTAAAAAATTAATTAATATATAATTAATATTTACTTTTATTTAAGTTT 840

QY 841 aattgttaatttggactatgtattatatttactatattttaaattgtttatagat 900  
DB 841 AATTCCTTAATTTTGGACTATGTATTTATTTCTACTATGTTTAAATGTTTATATAGAT 900  
QY 901 agtttaagaataataagtaatagttagtagagttttagagtgttaaccataaacaac 960  
DB 901 AGTTTAAAGTAATTAATTAATAGTATAGTAGAGTGTATAGAGTGTACCTTAACATAAC 960  
QY 961 tataacattatagtgagactaaatttcaataatttcttaattgtcttacccttttctgt 1020  
DB 961 TATAACATTTATGGGTGACTAATTTTCAATATATTTCTTATTTGCTTTCTTGTGT 1020  
QY 1021 atgttaagtcgtaataagtaataagttggttgccatggcactcgtgtcttggtttc 1080  
DB 1021 ATGTATGTCGGAATCTAATTAATTAATTAATGAGTGTGCTCATGCGACTCTGTGTTC 1080  
QY 1081 atgcatgggtcttgccgcaagaagaagaacagaacagaacagaacagaacagaaga 1140  
DB 1081 ATGCATGGGTCTTGCGCAAGAAAAGACAAAGAACAAAGAAAAGACAAACAGACAGA 1140  
QY 1141 caaaacgaacatcaacaacacaaactcaaatagtaacgtgtgataagatgcccgtcc 1200  
DB 1141 CAAAACGCAATCACAACAACTCAAAATTAATCACTGGCTGATCAAGATCGCGCTCC 1200  
QY 1201 atgtatgtctaaatgcataagcaagaacagtgcttaacatgacatttaagtctaac 1260  
DB 1201 ATGTATGTCCTAATGTCATGCAAGAACACAGCTGCTTATACATGCACTTAAATGGCTCAC 1260  
QY 1261 ccattctcaaccac 1320  
DB 1261 CCATCTCAACCCACCTGT 1320  
QY 1321 atataatcaatctctccgcacactcaaatcttcaactcaacaacagtcacactgcata 1380  
DB 1321 ATATATTTCAATCTCTCCGCGCACCTCAATTTCTTCACTTCAACACACGCAACCTGCATA 1380  
QY 1381 tgcgtgtatcccatggcccaaaatctccatgacatgtttccaaacacacttctctatata 1440  
DB 1381 TCGGCTCATCCCATGCCCCAAATCTCATGATGATGTTCCAAACACCTTCTCTTATATATA 1440  
QY 1441 taactataaactcttaataataactcaacttcttcaatcaatcaatcaatcaatcaat 1500  
DB 1441 TACCTTAATTAATACCTCTATATATATATATATATATATATATATATATATATATAT 1500  
QY 1501 ctacttactactataataatcccaacacacacacacacacacacacacacacacac 1560  
DB 1501 CTACTCTACTACTAT 1560  
QY 1561 tccctaaagcttcccttct 1620  
DB 1561 TCCCTTAAGCTCTTCCCTTTCTACGCTTTCCTTTGTTGGTCAATACTTCGTTGCTGTTA 1620  
QY 1621 ctcaacgtctgtagatcacccgcatctctctctcaaaagtaagtcctcgtlaaggcgc 1680  
DB 1621 CTCACGCTGCTGATCACCCGCAATCTCTCTACAAAGTAAGTCTCCGTAAGGCGC 1680  
QY 1681 tgaaggaacatggaactctcagaagactcttgcagaacaaacagatctggatcaagaaga 1740  
DB 1681 TGAAGGAACATGACACTCTTGAAGAGACTTCTTGAGAAACAAACGATGAGCATAGAGAGA 1740  
QY 1741 agtactccggcttcggttgaagttgtagtgtagtgtagtgtagtgtagtgtagtgtag 1800  
DB 1741 AGTACTCCGGCTTCGGTGAAGTTGCTAGAGTGCCACTTTCACCAACTTCTGATAGTCAAT 1800  
QY 1801 acttggagaagatctacccctcogaaccccgctcctaagaagtccaagttctcttctgatactg 1860  
DB 1801 ACTTTGGGAAGATCTACCTCGGAACCCCGCTCAAGAGTTCAACGTTCTTGTGATATG 1860  
QY 1861 gtccctctgactctggttccctctatctactgcaagagcaatgctctgcaagaacccac 1920  
DB 1861 GTTCTCTGACTCTTGCGTTCCTCTATCTACAGCAAGAGCAATGCGTGAAGAACACACC 1920  
QY 1921 aaagatcgatccgagaagaagtcgtccacacttccagaagaactgaagcaacctgtctatac 1980



|||||  
Db 1921 AAGATTCCATCGAGAAAGTGTCCACCTTCAGAACTTAGCAAACTTGTATAC 1980  
Oy 1981 actaagtgtaagtagatgacgaagaaattcttaggctatgataaccgtcactgtctccaaca 2040  
Db 1981 ACTAGCGTACAGGTGACGACGCAAGGAATCTTAGCTATATACCGTACTGTCTCCAA 2040  
Oy 2041 ttgtgacatcaacagacagtaggaacttagacccaagaaccaggtgagtctcact 2100  
Db 2041 TTGTGACATTTACACGACAGTAGTAGACTTAGACCCCAAGAACCGAGTGTACTTACCT 2100  
Oy 2101 atgcgaattcgaatggacatcccttgataggaatacccatcgtcgcgtgaagtactcga 2160  
Db 2101 AAGCAAAATTCATGGCATCTGTGTATGGCATACCATCGCTCGGTCAGAGTACTCGA 2160  
Oy 2161 taccgtgtttgaacaaatgatagaacacgaacctagtaactcaagaactgtctcgtt 2220  
Db 2161 TACCTGTGTTTACAAACAGATGAACCGACACCTAGTACTCAAGACTGTCTCGGTTT 2220  
Oy 2221 acatggacaggaatggccaggaagacatgctcacgcttggagctatgatacctact 2280  
Db 2221 ACATGACACGGAATGGCCAGAGAGCATGCTCACGCTTGAGCTATGTATCCATCTTACT 2280  
Oy 2281 acacagatctctcacttgggtccagtcactgtgcaagcagtagtctgcaatcactgtg 2340  
Db 2281 ACACAGATCTCTCAGTCTGGTTCCAGTCCAGTGCACGACGATGCGCAATTCAGTGTG 2340  
Oy 2341 acaagtcaacacacacgcggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2400  
Db 2341 ACAGTGTCCACATCAGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2400  
Oy 2401 ataccgtagtcgccaagctgtgcgagccatagacagcagatcttccaactcagcaagcta 2460  
Db 2401 ATACGGTATCGTCCAAAGCTGTGTGAGACTTAGACGACACTTCTCAACATTCAGCAAGCTA 2460  
Oy 2461 ttgagccaacacagaaacagtagttagttgaacatagatctgcgaacaaacttagtaca 2520  
Db 2461 TTGGAGCCACACAGAAACAGTAGAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2520  
Oy 2521 tggccaagttgtctttagatcaacgcaagatgtaaccactgacccctccgctata 2580  
Db 2521 TGGCTACAGTGTCTTTGATGATCAACGGCAAGATGTACCCACTGACCCCTCCGCTATA 2580  
Oy 2581 ccaagcaggaatcaagggttctgcacacagtgatctccaagtagaagcaatcccaagaat 2640  
Db 2581 CCAGCCAGAGATCAAGGTTCTGCACACAGTAGTCCAGAGTGAAGAACCTTCCCAAAAT 2640  
Oy 2641 ggaacttgggagatgtgttcaatcgtgactacaagcgtcttgaagggccaacaacc 2700  
Db 2641 GGAATCTTGGAGATGTGTTCATTCGTGATCTACAGCGTCTTTTGACAGGGCCAAACAC 2700  
Oy 2701 tctgtgggtctagctaaagcaatctgaagcttaagtaagtaagtaagtaagtaagtaagta 2760  
Db 2701 TCGTGGGCTAGCTAAAGCAATCTGAAGCTTAATGAATGAATGAATGAATGAATGAATGA 2760  
Oy 2761 gtaagagctcatgagagacatgaaatattgtatccgaacatgtaacagratataactga 2820  
Db 2761 GTAAGAGCTCATGAGAGATGGAATATTGTATCCGACCATGTAAACAGATATAAATCTGA 2820  
Oy 2821 gctcatctcactcttctatagaataaacaagaagtgtatagatatataactctatc 2880  
Db 2821 GCTCATCTCATCTTCTATGAATTAACAAGATGTATATATTAACACACTCTATC 2880  
Oy 2881 tatgacacttatgttctatagtaaaattccctctatatattataactctgtaactcgt 2940  
Db 2881 TATGACACCTTATTTGTATGATAAATTTCTCTTATTTATTAATCATCTGAATCGTG 2940  
Oy 2941 acggtctatgagatgcttcaaatagtaacaaacaagtgtgactataagactttctaa 3000  
Db 2941 ACGGCTTATGGAATGCTTCAATATAGTACAAAAACAATGTGACTATAGACTTTCTTAA 3000  
Oy 3001 caattctaacttagcatgtgaaacgagacataaagtgttaagaagacataaactataa 3060  
|||||

Db 3001 CAATCTAAGTATAGCATTTGTGAACGACATAGTGTTAAGAAGACATTAACATTTAA 3060  
Oy 3061 tggaaagaattgtctcccatlatalatalatalataaccactatgatatattaaga 3120  
Db 3061 TGGAAAGAATTGTCTCCATTTATTAATTAATTAATTAATTAATTAATTAATTAATTA 3120  
Oy 3121 tgttaagagacataacaattataaagaagaagtttgatccattatataatataatc 3180  
Db 3121 TGTTAAGGACACTTAACATTAATTAAGAGAGAAAGTTGTATTCATTTATTAATTAATAC 3180  
Oy 3181 taccatltatalatalataactlataccactaltlaaigtctltatalaagttlgaatcatg 3240  
Db 3181 TACCATTTATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3240  
Oy 3241 atattctataatttagttgatatgatatgaagaagtactaattgaaactctctactc 3300  
Db 3241 ATATTTCTATATATTTATGTTGATATGTATATGAAGGGATATTAATTTGAACCTCTTACAC 3300  
Oy 3301 tgtataaagttgagatcaccttaagtggtcactatlaatttatctgtctctacagat 3360  
Db 3301 TGTATTAAGTTGATCATCTTAAGAGTGGGTATTTATTTATTTATTTATTTATTTATTTAT 3360  
Oy 3361 aaaaaaaatlaagattgtgttga taaataatlgaaagattlaaataataataata 3420  
Db 3361 AAAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3420  
Oy 3421 ataaataacatataatataatgatatataaattatlaataataacattatctataaaa 3480  
Db 3421 ATAAATTAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3480  
Oy 3481 agfaaatgtgtacataactatacaatcgtttagccttgctgagacactcctaattat 3540  
Db 3481 AGTAATTAATTTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3540  
Oy 3541 taaacgagagtaaacatatttgacttctgtgtatttaacaattatlaattacaacta 3600  
Db 3541 TAAACGAGAGTAAACATATTTGACTTTTGTGTTATTAACAATTAATTAATTAATTAATTA 3600  
Oy 3601 atgaaatttttttttttttttttcgcaagaagaataaatttaattgaaggagcaatgtgt 3660  
Db 3601 ATGAATAATTTTGT 3660  
Oy 3661 gtccaatccttataacaacaacttccacaggaagtcaggtcggggaacaacaaaaaac 3720  
Db 3661 GTCCAAATCCTTATTAACCAACTTCCACAGAGGTGACGTGGGGACACAAAAAACC 3720  
Oy 3721 aggcagggaaattttaaattgggtgtcctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3780  
Db 3721 AGGCAAGGAAATTTTATTTATTTGCGTGTCTTGTGTGTGTGTGTGTGTGTGTGTGTGT 3780  
Oy 3781 caactaacataacccttttagcagatagacatggttgaacctgtgttagcttctttaa 3840  
Db 3781 CACTACACATAACCCCTTTAGCAGTAGAGCAATGAGTGTGACCGGTGCTTACGCTCTTTTA 3840  
Oy 3841 tttaattttttatcaagcaagaataataaataaataaagacactcaaggaatttc 3900  
Db 3841 TTTTATTTTATTCAGCAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3900  
Oy 3901 aaccttatacaaaacccaacaaagaatttccctagcaccttacaactaaggtacc 3957  
Db 3901 AACCCTTATACAAAACCCCAAAAACAAAGTTTCTTACACCCCTTAACCAATGAAGTACC 3957  
  
RESULT 2  
PHVBCSP PHVBCSP 3502 bp DNA linear PLN 17-JUN-1998  
DEFINITION Phaseolus vulgaris beta-type phaseolin storage protein gene,  
complete cds.  
ACCESSION J01263 M13758  
VERSION J01263.1 GI:3228361  
KEYWORDS  
SOURCE Phaseolus vulgaris.  
ORGANISM Phaseolus vulgaris

Eukaryote: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Phaseolus.
1 (bases 1371 to 2450: 2466 to 3294: 3322 to 3502)
Slightom,J.L., Sun,S.S.M. and Hall,T.C.
Complete nucleotide sequence of a French bean storage protein gene:
phaseolin
Proc. Natl. Acad. Sci. U.S.A. 80, 1897-1901 (1983)
2 (bases 1 to 3502)
Doyle,J.J., Schuler,M.A., Godette,W.D., Zenger,V., Beachy,R.N. and
Slightom,J.L.
The glycosylated seed storage proteins of Glycine max and Phaseolus
vulgaris: Structural homologues of genes and proteins
J. Biol. Chem. 261, 9228-9238 (1986)
3 (bases 2921 to 3502)
Hall,T.
Direct Submission
Submitted (27-APR-1993) Biology, Texas A&M University, College
Station, TX 77843-3155, USA
4 (bases 1 to 3502)
Hall,T.
Direct Submission
Submitted (16-JUN-1998) Biology, Texas A&M University, College
Station, TX 77843-3155, USA
Nucleotide sequence updated by submitter
On Jun 16, 1998 this sequence version replaced gi:169322.
location/qualifiers
1..3502
/organism="Phaseolus vulgaris"
/db\_xref="taxon:3885"
/clone="AF-PPVPh3.8; PPVPh 177.4-E1.6"
/clone\_1lb="Charon 24A clone 177.4"
join(1348..1860,1933..2123,2212..2292,2417..2662,
2791..3049,3153..3370)
/gene="beta-ppsp"
/gene="beta-ppsp"
<1348..1860
/gene="beta-ppsp"
/number=1
join(1548..1860,1933..2123,2212..2292,2417..2662,
2791..3049,3153..3370)
/gene="beta-ppsp"
/codon\_start=1
/product="beta-type phaseolin storage protein"
/protein\_id="AAC23610.1"
/db\_xref="GI:169323"
/translation="MMRARVPLILIGLIFLASISASFATSLREEESODNPFYFNSDN
SWRTFKNQYGHIRYLORFDOOSKRLQNLDEYRIVERPSKRETLILPQOAAELLIV
RSGSAILLVKRPDDRREYFLSDNPITSDHOKIPAGITFLVNDPEKEDLITOLAM
PVNPNQIHIEFLSTEAQSYLQESKILILASFSKTEELINRYLFEESGQEEQOE
GVIWNIDSEQIKELSKHAKSSRSKLSKODMTIGNEFNLERNVLSNIVSSIEME
EGALFVPHYYSKAIYILVANGEGAHVELVQNGKNEETLEYSEYRAELSKDVFVPA
YPAVKATSNVNTGFGINANNNNRNLGAGTDMVTSIGRALDGKDLGLFSSGSD
EVMKLINGSSTFYVDAAHHQEOQKSGHOEQKGRKGAIVY"
1548..1619
/gene="beta-ppsp"
join(1620..1860,1933..2123,2212..2292,2417..2662,
2791..3049,3153..3367)
/gene="beta-ppsp"
/product="beta-type phaseolin storage protein"
1861..1932
/gene="beta-ppsp"
/number=1
1933..2123
/gene="beta-ppsp"
/number=2
2124..2211
/gene="beta-ppsp"
/number=2
2212..2292
/gene="beta-ppsp"

/number=3
2293..2416
/gene="beta-ppsp"
/number=3
2417..2662
/gene="beta-ppsp"
/number=4
2663..2790
/gene="beta-ppsp"
/number=5
2791..3049
/gene="beta-ppsp"
/number=5
3050..3152
/gene="beta-ppsp"
/number=5
3153..3370
/gene="beta-ppsp"
/number=6
BASE COUNT 1124 a 678 c 572 g 1128 t
ORIGIN
Query Match 39.2%; Score 1551; DB 8; Length 3502;
Best Local Similarity 100.0%; Pred. No. 2.9e-235;
Matches 1551; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
7 gaatcatgtgtaaccgcatcatatagtgaaagtttgctctccgaggtgtt 66
|||||
1 GAATTCATTGACCTCCGATFATCATATGTAAGTTTGCTCTCGCCGGTGTT 60
|||||
67 tttaacctaatlaagggtttccacacaaatctggtatcatctcaattact 126
|||||
61 TTACTCTGATTAAAGGGTTTCCACCTAAATTCGTGATCATCTGACTTACT 120
|||||
127 gtactctaattctcataatcttgggtgtaaatlatacagctccgcacagatacc 186
|||||
121 GTTACTTTAATTTCTCAATAATCTTGGTGAATATACGCTTCGACAGATATCC 180
|||||
187 tcaaatattatattgtaaacatttcaacgcgataaaattttagaagtcgcgtc 246
|||||
181 TACAATTTATTATTGTTAAACATTTCAAACCCGATPAAATTTATGAGTCCGCT 240
|||||
247 accctaatgtagcttaacatttccatctgaataataatcttaatttagcgt 306
|||||
241 ACTTTAATGATGTCTAATATTTCATATGTAATATTAATTTAGCGTT 300
|||||
307 ggtagaagacataaagatttcttctatctctccatataaagtcttaatacaatat 366
|||||
301 GGTAGAAACATTAAGATTATTCTTATCTTCATATTAATGTTTAAATAAT 360
|||||
367 aacacaattcttaccttaagaagattcccatltaattlaaaaaataattatca 426
|||||
361 AACAATCTTTACCTTAAGAAAGATTTCCATTTTATTTTAAATATATTATCA 420
|||||
427 aatattttccaacacgtaaatctcataataaagttgtttcaaaagtaataaattla 486
|||||
421 AATATTTTCAACACGTAATCTCATATATTAAGTTGTTTCAAAAGTAATAATT 480
|||||
487 actcataatttcttctatctgacgtatcttaagcaacaacacgagtaacactgcat 546
|||||
481 ACTCCATTAATTTTATTCGACGTATCTTAAGCAACCCAGTGACCAACTGACAT 540
|||||
547 tttttcttgataaaaaaacataatcatatgatttttttataaataagaattt 606
|||||
541 TTTTTCCTTGATAAAAAATCAATTAATGATGATTTTATTAACAAATAATT 600
|||||
607 caaccaacaatcatgtgtgtaattctgaagcaagtcagtatacgaatctcataat 666
|||||
601 CACCAACAATCATTTGTGTGATTTCTGAGCAAGCATGATTAAGAAATCTAAT 660

REFERENCE	Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae									
AUTHORS	1 (sites)									
TITLE	Angenon,G., de Jaeger,G., Goossens,A. and Depicker,A.									
JOURNAL	Heterologous gene expression in plants									
FEATURES	Patent: WO 020899-A 5 03-JAN-2002;									
source	Vlaams Interuniversitair Instituut voor Biotechnologie vzw. (BE)									
BASE COUNT	1.									
ORIGIN	location/qualifiers									
	/organism="Phaseolus vulgaris"									
	/db_xref="taxon:3885"									
Query Match	34.2%; Score 135.6; DB 6; Length 1415;									
Best Local Similarity	99.6%; Pred. No. 4.3e-204;									
Matches 1367; Conservative	0; Mismatches 4; Indels 1; Gaps 1.									
QY	8	aattcatgttactccagatcatcatatagtgaagatttgctctctcgcgvggtttt	67							
Db	44	AATTATTGTACTCCCAATCATATATATGTGAATCTTTGGCTCTCTGCCGCGTGGTTTT	103							
QY	68	ttacctctattaaagggtttccacctaaatctcgttactcacttactctg	127							
Db	104	TTACCTCTATTAAAGGGGTTTTCCACCATAAAATTCGGTATTCATCTCCACTTACTTG	163							
QY	128	ttactttaatcttcataatcttggttgaaattacacgcttcgcacacatccct	187							
Db	164	TTACTTAAATTTCTCATATATCTTTGGTATTCGAATATTCACGCTTCGCGACAGATATCCCT	223							
QY	188	acaaattattatttgtttaaactttccaacgcgacataaaatttgaagtcgcgtcta	247							
Db	224	ACAAATTTATTATTTGTTAAACATTTTCAACCCGATTAATTTATGAAGCCGCTGA	283							
QY	248	tctttaatgtgtctaacattttcatatgtgaatatataatttacttaattttagcgttg	307							
Db	284	TCTTTAAATGTAGTCTCAACATTTTCATATTTGAATATTAATTTACTTAAATTTTACGCTTG	343							
QY	308	gtagaagaacataaagattatcttactctcttcctacataaagtgttaataacata	367							
Db	344	GTAGAAGACATATATGATTATTTATTTCTTCTTCATATTAATGTTTATATACATATTA	403							
QY	368	aacaaattcttaccttaagaagattcccacttlatatattttaaanaatalattacaa	427							
Db	404	AACAAATCTTTACTCTTAAGAAGATTTCCATTTTATATTTTAAATAATATTTATTCOA	463							
QY	428	ataatttccaacacgttaaatctcataataaagtgtgttccaagaataaataa	487							
Db	464	ATATTTTTCACACCGATTAATCTCATATTAATTAAGTTGTTTCAAAAGTAAATTTTAA	523							
QY	488	ctccataattttttatttcgcactgaactttaaagaaacccagtgacaaactaacatt	547							
Db	524	CTCCATTAATTTTATTTCGACTGATCTTTAAAGCAACCCGATCACCACTATCCATT	583							
QY	548	ttttctcttgataaaaaaacccaattlcatctgtatatttttllatacaatgaanaattlc	607							
Db	584	TTTTCTTTGGATAAAAAATTCATATTCATATGATTTTTTTTTTATACATGAATAATTTC	643							
QY	608	accacaaacatcatctgtgtatcttcgtgaagcaagtcagtltatgcanaattcatalaattc	667							
Db	644	ACCAAAACATATTTGTGTGATTTCTGGAAGCAAGCATGATATGCAAAATCTATAAATTC	703							
QY	668	ccatttgacactagaagatgaactgaagaatcgcgttttaactgagagaacatcttcaa	727							
Db	704	CCATTGTGACTAGGAGATGAAGATCGCTTTTACATGCGAGACACTTCTTCTPA	763							
QY	728	agtaatttataatagtactatataccaagaattcatalatcaaatcacaattact	787							
Db	764	AGTATTTTATTAATAGTTACTATATTCGAAGTTTCATATATCAATATCTCATATTTACT	823							
QY	788	tctaaaaaatatagataataaataatacttttttaattttaaagtttaagttt	847							

	BASE COUNT	299 a	308 c	262 g	304 t
Db	824	TCTAAAAAATTAATTAGTATTAATTAATAATTTACTTTTTTAATTTAAAGTTAACTTCGT	883		
Qy	848	gaattgtgactatgatattacatactactagtgtaaattgtttatagatagttaa	907		
Db	884	GAAATTTGACTATATGATTTATTTATTTACTACATGTTTAAATGGTTTATAGATAGTTTAA	943		
Qy	908	agtaataataagtaatgtagagagtgtagagtgtaaccctaactaaactaaca	967		
Db	944	AGTAATATTAAGTAATVAGTAGAGATGTGTAGAGTGTTACCCTAAACCAATAACTATAACA	1003		
Qy	968	ttaatggtggactaatcttcatalatatcttcatgtcctttaacctcttccttgtatgaag	1027		
Db	1004	TTTATAGTGAGACTATTTTCTCATATTTCTTCTTATTTCTTTTACTTTTCTTGATAGTAAAG	1063		
Qy	1028	tcgcgaactagaatcacagtggtgtgcatabgccactctgtggtccttttgtlcalgcatg	1087		
Db	1064	TCCGTAACCTAGAATTTACTGTGGTGTGCATGGCAGCACTGTGGTCTTTGGTTGATGACATG	1123		
Qy	1088	ggt-cttcgcgaagaaaagcaagaacaaaaaagaacaaaacgagagaaac	1146		
Db	1124	GATGCTTCGCCGAAGAAAAGACAAAAGAACAAAAGAAAAGACAGAGACAAAAAC	1183		
Qy	1147	gcaatcacacaaccacactcaaatltagtcacigtgcgtgatccaagatcgccgcgtcatagtat	1206		
Db	1184	GCAATTCACACACACACTCAANTTATGTACTGTGCGTGCATCAAGATGCGCGCTCCATGTAT	1243		
Qy	1207	gtctaataatgcataagcaaacacgltgcttaacatlgcactttaaitygtcaccacatc	1266		
Db	1244	GCTTAATATGCCATGCMAAAGCACACGTGCTPAAACATGCACCTTAATATGGCTCACCCATCT	1303		
Qy	1267	caaccacaacacaacacattgaccttttcttcatacatcaaccaaacacactgttatata	1326		
Db	1304	CACCCACACACAMAAACATTCCTTTTCTTCATCATCATCACCAACCAACCGTATATAT	1363		
Qy	1327	tcaatctcttcgccacactcaattcttcacttcaacaacagtcacactgca	1378		
Db	1364	JCATTCCTCTCCGCCACCTCAATTTCTTCACCTTCACACACAGTCACACTGCA	1415		
RESULT	4				
LOCUS	AX088019	1173 bp	DNA	linear	PAT 17-MAR-2001
DEFINITION	Sequence 1 from Patent WO0114571.				
ACCESSION	AX088019				
VERSION	AX088019.1	GI:13396947			
KEYWORDS	.				
SOURCE	Bos taurus				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bos.				
REFERENCE	1 (bases 1 to 1173)				
AUTHORS	van Rooijen G., Keon R.G., Boothe J. and Shen Y.				
TITLE	Commercial production of chymosin in plants				
JOURNAL	Patent: WO 0114571-A 1 01-MAR-2001;				
	Sembiosys Genetics Inc. (CA)				
FEATURES	location/Qualifiers				
source	1..1173				
	/organism="Bos taurus"				
	/db_xref="taxon:9913"				
	1..1173				
CDS	/note="unnamed protein product"				
	/codon_start=1				
	/protein_id="CAC34680.1"				
	/db_xref="gi:13396948"				
	/translation="MNFLKSPFYFAFLCFGOYFVAVTHAEITRIPLYKGSKLRAKLM EHGLIEDPLIQOYGISRSXGFGEVAVSLPTNLDSQFEGKIYLGPPORFTLFDYLP GSSDFWVSITYCKSNACKNHORPKSTPONLQKPLSHIGTGMQGLIGDTVTYV SNIDIOOTVGLSTOEDEPVTAEPEDILIMAPVASASESYIPEDMMNRHLYAOE LEFSYMRNGOESMLTLGADIDPSYTGSLHWNPVTVCOYMOPEDSVTISGVVUACE GCQAILDITGTSKLVGPSDDIINTQAIATATONOGEPDIDCNLSYMPFTVFETNGKK YPLPSPATTSODSGECITSGFSESNHSQMILGDVFIREYYSVPFRANLVGLAKAI"				

ORIGIN		29.6%; Score 1173; DB 6; Length 1173;
	Query Match	Best Local Similarity 100.0%; Pred. No. 1,3-175;
	Matches 1173; Conservative 0; Mismatches 0; Indels 0; Gaps	
OY	1554 atgaactctctaagttcttcccttccttaagcttcttcgttgttgycgcaactcgtt	1613
Db	1 ATGAACCTTCCTTAAGCTCCTTCCCTTTCCATCGGCTTCCCTTGTTGGTCAACTTCGT	60
OY	1614 gcttataacacgtgtctgatcalcacoccatccctctatcaaaagttaagtcacct	1673
Db	61 GCTGTACTACGCTGCTGAGATCACCCGATTCCCTCTAACAAGTAAGCTCTCCGT	120
OY	1674 aaggcgctgaaagacalgtgactctagaagactcttgcagaaaacaaglatygalc	1733
Db	121 AAGGGCTGAAGAACATGAGACTTTAGAAGACTTCTTGCAAAAACAAGATGACATC	180
OY	1734 agcagcaagatctccggcttcggttgaagttgtcagagtgatccattaccaactctgat	1793
Db	181 AGCAGCAAGTACTCCGCCCTTCGGTGAAGTTGGTAGGTCGCCTTACCACACTACCTTGAT	240
OY	1794 agtcaaatcttggaaagatctaacctcccgaaaccccgcacaagtltaaccgttctt	1853
Db	241 AGTCAATACTTTGGGAAGATCTACTCTGGAAACCCCCTCAAGAGTTCAACGTTCTT	300
OY	1854 gatactgttcccttgactctgtggttcccttatctacttgaagaagcaatltcgtcaag	1913
Db	301 GATACCTGGTTCCTCGACTTCCTGGTTCCCTCATCTACTGCAAGCAATCCTGCAAG	360
OY	1914 aaccaccaagaatcgcataccgagaagaatgctccacttccagaacttaagcaaccttg	1973
Db	361 AACCCCAAAMATTGATTCGATCCGAGAAAGTCGTCCACTTCCAAACACTTAGCCAACCCTTG	420
OY	1974 tctataacatacaggtataaggtatgacatgacagaagacttagggctatgataccgttaact	2033
Db	421 TCTATACACTAGGTTACAGGTAGCATGCAAGAACTTTAGGCTATGATACGCTCACTGTC	480
OY	2034 locaaacattgtgagcatlcaacagacagtagagcttagaacccaagaacagatgatl	2093
Db	481 TCCAACATTGTGGACATTCAACAAGACAGTAGAGACTTAGCAACCAAGAACAGAGTAGTGTC	540
OY	2094 ttcaactatgcagaatcgcatacgatcccttgltatggaalaccactcgctcgctcagag	2153
Db	541 TTCACCTATGCAGAAATTCGATGGCATTCCTTGATGGCATACCATCCGCTCCGTCAGAG	600
OY	2154 tacttgataactgtgtttgacaacatgatgaacagcacctagtatgataaactgttc	2213
Db	601 TACTCGATACCTGTTGTTGCAACATGATATAAACCGACCCCTAGTACTCAABACTTGTC	660
OY	2214 lcggtttcaacygacagaatlggccagagaagatgctcaagcttbgagctatltatcca	2273
Db	661 TGCGTTTCAATGGACAGAGAAATGGCGAAGAGACATGCTCAAGCTTGAGACTTTGATCCA	720
OY	2274 tctactatacacaagatctcttcaacttgggttccagtactacttbtgacagtaactygaactlc	2333
Db	721 TCTCATCTACACAGAGATCTCTCACTGGGTTCCAGTCACGTGTGACGAGCACTAGCAATTTC	780
OY	2334 acctgtgacagtgatccaccatcaacgcgtgtgtgttgcatactgtgaagtygatgtcaagct	2393
Db	781 ACTGTGGACACTGTACACATTCAGCGGCTGTGGTTGTTGCAATGTGAAGGTGAAGTCAAGCT	840
OY	2394 atcttggatlaacggttactgcataacgcttggctggagccccagagagaaactttaacaattag	2453
Db	841 ATCTTGGATACCGGTAGCTCCAACTGCTGTGGACCTTAGACGACCACTTTCTCAACATTTCAG	900
OY	2454 caagctatctgagccacacagaaccaaactiaagtgtagtttgaacatalagatctgcgaacaactt	2513
Db	901 CAAGCTATTGGAGGCCACACAGAAACCAATAGGRTAGGATTTGACATPATATTGGCAACAACGTT	960
OY	2514 agctacatgctactagttgtctttagatacaacgcaagatgtaccacttgaccccttc	2573

Db 961 AGCTACAGCCTACAGTGTCTTGAGATCAACGGCAAGATGTACCACTGACCCCTCC 1020  
Qy 2574 gacctacagcagcagatcaagggtctgcaccacagtgattccagagatgagaaacattcc 2633  
Db 1021 GCGTATACACCGACGATCAAGGGTCTGACACAGTGCATTCACAGATGAGAACCATTC 1080  
Qy 2634 cagaatagatcttggagatgtgttcattcgttgagtaactacaagcttcttgacagggcc 2693  
Db 1081 CAGAAATGATCTGGGAGATGTGTTCATGTGTAGTACTACAGCGCTCTTGACAGGGCC 1140  
Qy 2694 aacaacctcgttgggctagctaaagaactcga 2726  
Db 1141 AACAACTCGTTGGCGTAGCTAAACCAATCTGA 1173  
RESULT 5  
AX252300  
LOCUS AX252300 2970 bp DNA linear PAT 05-OCT-2001  
DEFINITION Sequence 2 from Patent WO0168887.  
ACCESSION AX252300  
VERSION AX252300.1 GI:15985641  
KEYWORDS  
SOURCE synthetic construct.  
ORGANISM synthetic construct  
REFERENCE 1 (bases 1 to 2970)  
AUTHORS Jung, R. and Kinney, A.J.  
TITLE Hypoallergenic transgenic soybeans  
JOURNAL Patent: WO 0168887-A 2 20-SEP-2001;  
E. I. du Pont de Nemours and Company (US) ; PIONEER HI-BRED  
INTERNATIONAL, INC. (US)  
FEATURES  
source 1. 2970  
Location/Qualifiers  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="Chimeric construct"  
BASE COUNT 1018 a 547 c 527 g 878 t  
ORIGIN  
Query Match 29.1%; Score 1149.8; DB 6; Length 2970;  
Best Local Similarity 99.0%; Pred. No. 5.4e-172;  
Matches 1157; Conservative 0; Mismatches 12; Indels 0; Gaps 0;  
Qy 2735 aagatgacacaaatgcatgtagtgaagagcctcagagatgagatattgtatc 2794  
Db 1796 AAGATGACAACTAAATGATGATAGGTGTAAGAGCTCATGAGAGCATGGAATATTGTATC 1855  
Qy 2795 cgaccatgtaacagtlataactagagctcatctcactctctctatgaataaacaag 2854  
Db 1856 CGACCATGTACAGTATATTAAGTGAAGCTCATCTCACTTCTTCATGATTAACAAAGG 1915  
Qy 2855 atgtatgatatataaaccctctatctgcaaccttatgttctatgataattcctc 2914  
Db 1916 ATGTATGATATATTAATTAACACCTATCTATGCAACCTATTGTCTCATGATAATTTCCCT 1975  
Qy 2915 tatattataaatcatcgtgaatcgtgaagcttatgagatccttcaaatagtaaaaaac 2974  
Db 1976 TATATTTATTAATCATCTGGAATCGTGAAGGCTTATGGAATGCTTCAATATGACAAAAAC 2035  
Qy 2975 aaatgtactataaagacttcttaacaacttcaacttgaatgtgaagagacataa 3034  
Db 2036 AAATGTGATATTAAGACTTCTTAACAATCTTAACCTTAGATGTGAAGAGACATTA 2095  
Qy 3035 gtgttaagaagaacataaatataatgagaagattgtctcaattatataatata 3094  
Db 2096 GTGTAAACAAACATTAACAAATTAATGAAGAACTTGTCTCCATTTATATATATATA 2155  
Qy 3095 ttaccacttatgtattatattagtagtgaagagacataaacaattaaagaagaag 3154  
Db 2156 TTACCCACTTATGTATTTATTTAGATTTTAAGGAGACATTAACATTTTAAGAGAGAG 2215  
Qy 3155 ttgtatccattatataataataactacattatataattatacttaccactatt 3214

Db 2216 TTTGTATCCATTATATATATATATATACACCCATTATATATATATATATATATATAT 2275  
Qy 3215 aatgctttataagtttgatccatgatattcttaataatttaagtgtatgatataatga 3274  
Db 2276 AATGTCTTATTAAGGTTTGATCAGATATATTTCTAATATTTAGTTGATATATATGA 2335  
Qy 3275 agggtaactattgaactcctctactctgtataaagtttgagatcccttaagttgct 3334  
Db 2336 AGGTACTATTGAACTCTTACTCTGTATTAAGGTTTGATCATTAAAGTGGGCT 2395  
Qy 3335 attaatattatgtcttcttaacagataaaaaaaatattgattgtgttgaataaat 3394  
Db 2396 ATTTATTTTATTTGCTTCTTACAGATTAATAAATAATATGATGTTGTTGATAAATAT 2455  
Qy 3395 tgaagattttaaataataataataataataacataataatgatatataatattat 3454  
Db 2456 TGAAGGATTTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 2515  
Qy 3455 tataataacattatctataaagaatgaatattgtcatataactctataacatgctta 3514  
Db 2516 TATATATATACATTTATCTATTAATAAAGTAAATATTGTCATTAATCTATACATGCTTTA 2575  
Qy 3515 gcttgcctggagcactcctaattttaaacgagatgaacataattgactttgttga 3574  
Db 2576 GCGTGTGAGAGACTCTCAATTTATTAACGAGTAAACATATTGACTTTTGGTTA 2635  
Qy 3575 tttaacaattatatttaacacataatgaatttttttttctcggcaaggaataa 3634  
Db 2636 TTTAACAAATATATTTTAAACATATGAATTTTATTTTATTTTATTTTATTTTATTTA 2695  
Qy 3635 aattaattagagaggaatattgtgtcccaatccttataaaccacttccagaaga 3694  
Db 2696 AATTAAATTTAAAGAGACATGCTGTGTCACATCTTATTAACACATCTTCAACAGAA 2755  
Qy 3695 gttcagatcgggagacaaataaaacaaagcagggaaatttttlaatttggttctcg 3754  
Db 2756 AGTCAAGTCAGAGACACAAAAAACAAGCAAGAAATTTTAAATTTGATTTGCTTGG 2815  
Qy 3755 ttgtctgataattatgacagtaataaacatcacataaccctttagcagtagagaatg 3814  
Db 2816 TTTTGTGATATATTTATTCAGATTAACATCACTACATCACTTTTATGAGCATGATG 2875  
Qy 3815 gttgacgttgcttagctctcttattatttttttttctacagcaagaataaataat 3874  
Db 2876 GTTGACCGTGTCTTAGCTCTTTTATTTTATTTTATTTTATTTATTCAGCAAGATTAAT 2935  
Qy 3875 aaatlgagacactcagagatgtttcaac 3903  
Db 2936 AAATGAGACACTTCAAGGATGTTTCAAC 2964  
RESULT 6  
BOVCHYMOB  
LOCUS BOVCHYMOB 1305 bp mRNA linear MAM 26-APR-1993  
DEFINITION bovine chymosin b (rennin) mRNA.  
ACCESSION J000003  
VERSION J000003.1 GI:162859  
KEYWORDS chymosin; chymosin B; protease; rennin.  
SOURCE bovine (calf) cdna of fourth stomach mucosa mRNA.  
ORGANISM Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovidae; Bovinae; Bos.  
REFERENCE 1 (bases 1 to 1305)  
AUTHORS Harris, T.J.R., Lowe, P.A., Lyons, A., Thomas, P.G., Eaton, M.A.W.,  
Milligan, T.A., Patel, T.P., Bose, C.C., Carey, N.H. and Doel, M.T.  
TITLE Molecular cloning and nucleotide sequence of cdna coding for calf  
preprochymosin  
JOURNAL Nucleic Acids Res. 10, 2177-2187 (1982)  
MEDLINE 82221400  
COMMENT chymosin is the major proteolytic enzyme in the fourth stomach of  
the unweaned calf. two chromatographically different forms, a and



Qy 1729 gcatcagcagcaagctacccgctcgttgaaagtctagcgtgcacttaccaactacc 1788  
 Db 220 GCATAGCAGCAAGTACTCCGGCTTCGGGGAGTGGCCAGCTGCCCTCGACCACTATAC 279  
 Qy 1789 ttgatagtcactactcttggaagatctccttcggaaccccgctcaagaagctcgttc 1848  
 Db 280 TGGAAAGTCACTACTTGGGAAGATCTACCTCGGAGACCCGCCGACGAGTTACACCGTCC 339  
 Qy 1849 tcttgatagctggttcctcctcctcctcgttcctctatctactgccaagagcaatgctt 1908  
 Db 340 TGTTTGAACTAGTGGCTCTCTGTGACTCTGGGTACCTCTATCTACTGCAAGCAATGCTT 399  
 Qy 1909 gcaagaaccacaaagatctgcatccgagaagctgccaacttcaggaactaggcaaac 1968  
 Db 400 GCAAAAACCAACAGCGCTTCGACCCGAGAAATGCTCCACTTCCAGAACTCGGGCAAC 459  
 Qy 1969 ccttgctactacactagcgttagcaggttagcagaagaaactcttaggtatgatacgttca 2028  
 Db 460 CCTGTCTATCCACTACACGAGGAGCAGCAGCATGCAAGGCTATGCTATGACACCGTCA 519  
 Qy 2029 ctgtctccaacattgtgacattcaacagagtagagacttagcaacccaagacagctg 2088  
 Db 520 CTGTCTCCAAATGTGTGGACATCCAGGAGACGTAGGCTGAGCACCAAGGAGCCGGGG 579  
 Qy 2089 atgtctcactatgcaagaattcgaatgcatcctgtgtatgtgcatccatcgtcgcgt 2148  
 Db 580 ACGTCTTACCTATGCGCAATTCGACGGGATCTGTGGGATGGCCTTACCCCTCGCTCGCT 639  
 Qy 2149 cagaatctcgaatacctgtgttgacaacatgataaacagcaactagtagctcaagact 2208  
 Db 640 CAGAGTACTCATACCCCTGTTGACAAATGATGAACAGGACCTGCTGGGCCCAAGACC 699  
 Qy 2209 tttctcgtgttatatgacagaggaatgacaggaagagcatctcaacgtttgagatattg 2268  
 Db 700 TGTCTCTCGGTTTAAATGACACAGAAATGGCCAGAGAGCATGCTCAGCTGGGGGCAATCG 759  
 Qy 2269 atccatcctactacagagatctcttcacttggttcacgtcactgttcagcagatctgac 2328  
 Db 760 ACCGCTCTACTACACAGAGGTCCTGCGACGTGGTCCGCTGACAGTGCAGCATGCTGCG 819  
 Qy 2329 aattcactggtgacaggtgtcaaccataagcgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2388  
 Db 820 AGTTCACTGTGACAGTGTACATCACACGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGT 879  
 Qy 2389 aagatctatgataccggtacgctcgaagctggttcgagcctagcagcagcatctccaca 2448  
 Db 880 AGGCAATCTCTGACACGGGCACTTCACAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 939  
 Qy 2449 ttcaagcaagctatgagcagcagcagcaagcagtagcgtgtgtgtgtgtgtgtgtgtgt 2508  
 Db 940 TCCAGCAGGCGCATTTGGAGCCACACAGAAACGATGAGTGTGTGTGTGTGTGTGTGTGT 999  
 Qy 2509 accttagctataatgctcactaggtgtctttagatcaacagcagaagtgtaccactgaccc 2568  
 Db 1000 ACCGACTATATCCCACTGT 1059  
 Qy 2569 ccttcgacatatacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 2628  
 Db 1060 CCTCGGCTATACACAGCAGCAAGAGGCTTCTGTACACAGTGTGTGTGTGTGTGTGTGTGT 1119  
 Qy 2629 attcccaagaatgtgactcttggaagatgtgttcattcgtgtgatacctaagcgtcttgaca 2688  
 Db 1120 ATTCCAGAAATGTGATCTGTGGGGATGTTTTCATCCGAGATATATTACACGCTTTTGACA 1179  
 Qy 2689 gggcgcaacaactcgtgtggtcgttagcctaagcaatcgaagctt 2731  
 Db 1180 GGGCGCAACACTCTGTGGGCTGTGGCCAAAGCCATCTGACTCTGT 1222

RESULT 8  
 E00042 E00042 1460 bp RNA Linear PAT 29-SEP-1997  
 LOCUS  
 DEFINITION DNA coding of pre-pro rennin.

ACCESSION E00042  
 VERSION E00042.1 GI:2168348  
 KEYWORDS JP 1982141287-A/1.  
 SOURCE unidentified.  
 ORGANISM unidentified.  
 REFERENCE 1 (bases 1 to 1460)  
 AUTHORS Baanadetsute,R.A., Jieni,M., Donarudo,T.M., Arison,T.R. and Jierarudo,F.B.  
 TITLE RENNIN, PREPRORENNIN OR PRORENNIN GENE OBTAINED FROM RECOMBINED DNA MATERIAL AND LIVE CELL CONTAINING GENE Patent: JP 1982141287-A 1 01-SEP-1982;  
 JOURNAL KOBUNSHI KAGAKU ZASSHI 89:103-107 (1982)  
 COMMENT OS calf  
 PN JP 1982141287-A/1  
 PD 01-SEP-1982  
 PF 14-JAN-1982 JP 1982003556  
 PR 16-JAN-1981 US 81 225717, 01-DEC-1981 US 81 325481 PI  
 BAANADETSUTE RABUTSUTSUKII ARUFUODO, JIENI MAO, PI DONARUDO  
 TEIRAA MOJIA  
 PI ARISON TAUNTON RIGUBIT, JIERARUDO FURANSHISU BUOBUSU PC  
 C12N1/00,C07G7/00,C07H21/04,C12N15/00,C12P21/00//C12N1/18, PC  
 C12N1/20,  
 PC C12P19/34,C12R1/19,C12R1/865;  
 CC strandedness: Double;  
 CC topology: linear;  
 CC hypothetical: No;  
 CC anti-sense: No;  
 CC \*source: tissue\_type=stomach; Location/Qualifiers  
 FH Key  
 FT CDS  
 FT 205..1350  
 FT /product=pre-pro rennin/  
 FT 205..1347  
 FT /product=pre-pro rennin/  
 FT 253..1350  
 FT /product=pro rennin/  
 FT 253..1347  
 FT /product=pro rennin/  
 FT 278..1350  
 FT /product=pro rennin/  
 FT 278..1347  
 FT /product=rennin/  
 FT mat\_peptide  
 FT 278..1347  
 FT /product=rennin/  
 FT FT  
 FEATURES  
 source Location/Qualifiers  
 1..1460  
 /organism=unidentified/  
 /db\_xref=taxon:32644  
 BASE COUNT 328 a 440 c 398 g 294 t  
 ORIGIN  
 Query Match 22.7%; Score 897.2; DB 6; Length 1460;  
 Best Local Similarity 87.7%; Pred. No. 3.9e-132;  
 Matches 980; Conservative 0; Mismatches 138; Indels 0; Gaps 0;



Db	473	TGTTTACACATGGCTCTGACTTCTGGGTACCCCTATCTACTGACAGCAATGCT	532
Qy	1909	gcaagacacacaaagattcgatccgaaagtcgctccacccacgaacttaggcaac	1968
Db	533	GCAAAACACACGCGCTTGACCCGAGAAAGTCGTCCACTTCCAGAACCTGGCGAAGC	592
Qy	1969	ccttgctctatacctaagtcagtagtaagcattgcaaggaatccttagctatagatcgta	2028
Db	593	CCCTGCTATACCACTACGGGACAGGCGCATGACAGGCACTCCGGGCTATGACACCTCA	652
Qy	2029	ctgttcccaacattgtgagatcccaacagtcagtaagtaagtaagcaaaccaagctg	2088
Db	653	CTGTCTCCACATTTGTGACATCCAGACAGATGAGGCTTGAGACACCAGAGCCCGGG	712
Qy	2089	atgtcttcacctaacycagaattcgatggcactccttgatggatacaccatcgctcgt	2148
Db	713	ACGCTTACACTATGGCGAATTTGACGGGATTCCTGGGATGGCTTACCCCTCGCTGCT	772
Qy	2149	cagaagtaactgataactgtgtttgacaacatgtaaacgcgaacactagtagctaaagt	2208
Db	773	CAGAGTACTGATACCCGCTTTGACAAATGATGAACAGGCACTGTGTGCCCCAAGACC	832
Qy	2209	tggttcgtgttacaatgagaagaatgacagagagacatgctcagcttgagatactg	2268
Db	833	TGTTCTCGTTTACATGAGACGGAATGGCCAGAGACATGCTTACCGTGGGGCCATCG	892
Qy	2269	atccatctactaacacagagatctcttcaactgtgttcacagtaactgtgcaagtaactg	2328
Db	893	ACCCGTCTACTACACAGAGGTCCTGCACGTGGTGGCCGTGACAGTGAAGCATGTCGC	952
Qy	2329	aattcaactgtgagacagtgtcaacatcagcaggtgtgtgtgtgtgtgtgtgtgtgtgt	2388
Db	953	AGTTCACGTGTGACAGTGTCAACATCAGCCGCTGTGTGTGTGTGTGTGTGTGTGTGT	1012
Qy	2389	aagatctcttgagacaggtgacgcgaagctgtgcgaacactgacagcagacattccaac	2448
Db	1013	AGGCACTCTGAGACACGGGCACTCTCAAGCTGTGGGCCCCAGACGACACTCTCAACA	1072
Qy	2449	ttcagcaagctatttggacacacagaaacagtaacgtgtgatttgcataagattgacga	2508
Db	1073	TCCAGACAGGCCATTGGAGCCACACAGAAACAGTACGATGATTTGACATTCAGTGC	1132
Qy	2509	accttaactatactgactacagtgctcttgtagatcaagaggaagatgtgacccactgac	2568
Db	1133	ACCTGAGCTACATGCCCACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	1192
Qy	2569	ccttcgactatacagcagagatcaaggtgttcgaacagctggaattccagagtgaac	2628
Db	1193	CTTCGCGCTATACAGGACGACAGGCGCTTCTGACCAAGTGGCTTCCAGATGAAAAATC	1252
Qy	2629	attccagaagaatgagatcttggagagatgtgtcatctcgttgatactacagcgtcttg	2688
Db	1253	ATTCCCAAGAAATGTGCTCTGGGGGATGTTTCATCCGAGAGATATACAGCGTCTTG	1312
Qy	2689	ggggcaacaacactcgttgggttagtaagaatctga	2726
Db	1313	GGGCCAACACACTCGTGGGCTGGCCAAAGCCATCTGA	1350
RESULT	9		
LOCUS	A15836	1291 bp	DNA
DEFINITION	chymosin gene.		linear
ACCESSION	A15836		
VERSION	A15836.1		
KEYWORDS	GI:488959		
SOURCE			
ORGANISM	Bos taurus		
REFERENCE	Eukaryota; Metazoa; Chordata; Cranialia; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;		
	Bovidae; Bovinae; Bos.		
	1 (bases 1 to 1291)		
	Simons, A.F.M. and De Vos, W.M.		

TITLE	DNA fragments, containing a lactic acid bacterium-specific regulator region for the expression of genes coding for normally heterologous proteins									
JOURNAL	Patent: EP 0307011-A 5 15-MAR-1989; NEDERLANDS INSTITUUT VOOR ZUIVELONDERZOEK									
FEATURES	Location/Qualifiers									
source	1. .1291									
BASE COUNT	323 a	382 c	328 g	257 t	1 others					
ORIGIN										
Query Match	22.6%;	Score 896.2;	DB 6;	Length 1291;						
Best Local Similarity	87.6%;	Pred. No. 5.7e-132;								
Matches 979; Conservative	0;	Mismatches 139;	Indels	0;	Gaps	0;				
Qy	1609	tcgttgcgttactcacacgctgctgagatcaaccgcatctctctacaagtgatgtc	1668							
Db	9	TCTTCGCTCTCTCCAGGGCGCTGATACACAGATCCCTGTACAAAGCAAGTCTC	68							
Qy	1669	tcgtaagcgctgaaagaaatggaactctagaagactcttgcaagaacaagatg	1728							
Db	69	TGAGGAAGCGCTGAAGAGCATGGCTTGTGAGGACTTCTCAGAAACAGCAGTATG	128							
Qy	1729	gcatacagagaagtaactcggcttggtgtaagtgtgtagctgtccacttaacatac	1788							
Db	129	GCATCAGCAGCAAGTACTCCGCTTGGGAGGTGGCCAGGTCCTCTGACCAACTAC	188							
Qy	1789	ttgataagtaacttgggaagaatcaactcggaaacccgcctcaagaagtcacgttc	1848							
Db	189	TGATATGCACTACTTTGGAGATCTACCTGGGACCCGCCAGAGATTCACCGTGC	248							
Qy	1849	tcttgabactgttctcctgtaactctggttccctctatactactgcaagaagtcct	1908							
Db	249	TGTTGACACTGGCTCCTGACTTGTGGTACCTCTATCTACTGCAAGCAATGCT	308							
Qy	1909	gcaagaacacacaaagtatgcatccgagaagtgctgcaccttcagaacttaggcaac	1968							
Db	309	GCAAAACACACAGCGCTTGCACCCGAGAAAGTGTCCACTTCCAGAACCTGGGCAAGC	368							
Qy	1969	ccttgctatacactagtcgaatgacatgcaagaagaatttagatgatacgtca	2028							
Db	369	CCCTGTCTATCCTACAGCAAGCAAGCAAGGCGATCCTGGGCTATGACACCGTCA	428							
Qy	2029	ctgtctcaacatgttgacatcaacacagcagtagaacttagcaacccaagaacagtg	2088							
Db	429	CTGTCTCCAACTGTGTGACATCCAGCAGACAGTAAAGCCTTGAAGCACCAGAGCCGGG	488							
Qy	2089	atgtcttcaactatgcaaatctgcatggtccttggtatgcataccatcgctcgt	2148							
Db	489	ACGTCTTACCTATGCCCAATTCGACGGGATCTGGGATGGCTTACCCCTCGCTCGCT	548							
Qy	2149	cagaatctcgaatcgtgttgcaacaatgatacgaacgaacttagttagctagaact	2208							
Db	549	CAGATCTATCGATACCGCTGTTTGAACAATGATGAACAGGACTGTGGGCCCAAGACC	608							
Qy	2209	tgcttcggttaacatgagcaaggaatggccagagagacatgctcaagcttgagactatg	2268							
Db	609	TGTTCTCGGTTTACATGAGACAGGAATGGCCAGGAAGCATGCTCAAGCTGGGGCCATCG	668							
Qy	2269	atccatctactaacacagatctcttcaactggttcagactcgttgcaagcagatagc	2328							
Db	669	ACCCNTCTACTACACAGGCTCCCTGCACTGGGTCCCGGTGACAGTGCAGCATGTCTGC	728							
Qy	2329	aattcaactgtgacagtgatcaacatcagcagtggtgtgtgtgtgtgtgtgtgtgtgt	2388							
Db	729	AGTTCACTGTGACAGTGTGACATCAGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGT	788							
Qy	2389	aagcatcttgataccggtatcgatcaagctgtgtgcaacttagcagcagatctccaac	2448							
Db	789	AGGCACTCTGAGACACAGGCACTCCAAGCTGTGTGGGCCAGCAGCAATCTCAACA	848							







Qy	2210	gtctcgtgttaccatgagcaggaaatgcccagagagacatgctcacgcttgagactatga	2269
Db	1269	gtttcgtgtttacatgacaggaatggccagagacatgctcacgcttgagagccatgca	1328
Qy	2270	tcatctaccatcaccaagatctcttcactggttccatgaactgtgcagcagtaactga	2329
Db	1339	ccccctctacacacaggggtccctgcattgggtgcccctgacagagtcacatctgcca	1388
Qy	2330	attcacgtgagcaggtgcacacatcagcgtgtgtgtgttgctgactgagagtgatga	2389
Db	1389	gttcactctgacacatgctcaccatcagcgtgtgtgtgtgtgtgtgtgtgtgtgtgt	1448
Qy	2390	agctactctgagatccggttacgtccaaagctgtgcagacctaagcagacatctcaaat	2449
Db	1449	ggccatctctgacacagcgcacacctccagctggtggtggtggtggtggtggtggtg	1508
Qy	2450	tcagaaagctttgagcagcacaacagacagcagtgagtgagttgacatagatgcgaca	2509
Db	1509	ccagcagcagcagtgagcagcagcagcagcagcagcagcagcagcagcagcagcag	1568
Qy	2510	ccttagctacatgctcctacatgctgtctgtgagatcaagcagcagcagcagcagcag	2569
Db	1569	ccttagctacatgctcctacatgctgtctgtgagatcaagcagcagcagcagcagcag	1628
Qy	2570	ctccgctactacacagcagcagcagcagcagcagcagcagcagcagcagcagcagcag	2629
Db	1629	ctccgctactacacagcagcagcagcagcagcagcagcagcagcagcagcagcagcag	1688
Qy	2630	ttcccaagaatgagatctgtgagagatgtgtcaatcgtgagtaactacagcgtcttgacg	2689
Db	1689	ttcccaagaatgagatctgtgagagatgtgtcaatcgtgagtaactacagcgtcttgacg	1748
Qy	2690	ggccaaacacactcgtgtgagcagcagcagcagcagcagcagcagcagcagcagcagc	2749
Db	1749	ggccaaacacactcgtgtgagcagcagcagcagcagcagcagcagcagcagcagcagc	1808
Qy	2750	tgcatgtgagtgagatgagcagcagcagcagcagcagcagcagcagcagcagcagc	2809
Db	1809	ccaagcgtttagtgagatgagcagcagcagcagcagcagcagcagcagcagcagcagc	1868
Qy	2810	ataatactgagcctccatctcaactctctctatgaata	2847
Db	1869	ttttatgactgagtgagcagcagcagcagcagcagcagcagcagcagcagcagcagc	1906
RESULT 13			
BOVCHYMOA			
LOCUS	1275 bp	mRNA	linear
DEFINITION	bovine chymosin a (rennin) mrna.		
ACCESSION	J00002		
VERSION	J00002.1	GI:162857	
KEYWORDS	chymosin; chymosin A; rennin.		
SOURCE	bovine (calf) cdna of fourth stomach mucosa mrna.		
ORGANISM	Bos taurus		
REFERENCE	Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi; Eukaryota; Metazoa; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovinae; Bos.		
AUTHORS	Molr,D., Mao,J.L., Schumm,J.W., Vovis,G.F., Alford,B.L. and Teunton-Rigby,A.		
TITLE	molecular cloning and characterization of double-stranded cdna coding for bovine chymosin		
JOURNAL	Gene 19, 127-138 (1982)		
MEDLINE	83054629		
COMMENT	chymosin (rennin) is the major proteolytic enzyme in the fourth stomach of the unweaned calf. two chromatographically distinct forms are known and a third seems likely (see bovychmob and bovychmoc). this sequence has been tentatively identified as chymosin a; it differs from chymosin b by only two amino acids and from chymosin c by another amino acid. [1] argues that the different chymosins are probably polymorphic variants of a single chymosin gene.		
FEATURES	Location/Qualifiers		

source	1. 1275	/organism="Bos taurus"	
CDS	21. 1166	/db_xref="taxon:9913"	
		/note="preprochymosin a"	
		/codon_start=1	
		/protein_id="AAA30447.1"	
		/db_xref="GI:162858"	
		/translation="MRCILVYLAVPALSGAEIIRIPLYKGSIRKALKKEHGLLEDFL OKOXYGSSSKSGFGEVAVPLTYLIDSQYRGKTYLGRPEPFVLPDTSDDWVPS IYCKSNACKNHQRFDPKRSSTFQNLGKPLSHYGTSQMOGLIGDTVTSNLIIDQOT VGLSTQEPGDVFTYAEFDGILGMAYPSLASEYSIPVEDNMNRHLVADLEPSYMDRNGOESMLTGLAIDPCYTGSLHWPVTVOOYQFVDSYVTSIGVVVACGGCOALIDTGSTKLVPSSDLINLOAIGATONQYDEEDIDCDMLNSVPTVFEINQKMYLPFSAYTSODGFCISGFSQENSHSKWILGDVPIREYISYVDNRANNLVGLAKAI"	
sig-peptide	24. .68	/note="presence"	
mat-peptide	195. .1163	/product="chymosin a"	
BASE COUNT	293 a 391 c 336 g 255 t		
ORIGIN	20 bases upstream from codon 1.		
Query Match	22.6%; Score 894; DB 4; Length 1275;		
Best Local Similarity	87.5%; Pred. No. 1.3e-131;		
Matches	978; Conservative 0; Mismatches 140; Indels 0; Gaps 0;		
Qy	1609	tcgtgtgtgttactcaagctgctgagatcacccgcatctctctacaagaagtgaatctc	1668
Db	49	tcttgcgtcttcccccagggcgtgagatcacccgcatctctctacaagaagtgaatctc	108
Qy	1669	tcgtaagcgcgtgaggaagacatgagactctgagaagactcttgagagaacaacagatg	1728
Db	109	tgagagaagcgttgaggaagacatgagactctgagaagactcttgagagaacaacagatg	168
Qy	1729	gcatacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag	1788
Db	169	gcatacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag	228
Qy	1789	ttgagatgaactactcttggaagatctactctggaaccccgctctaagaagtctacgcttc	1848
Db	229	tggatgacgagatgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc	288
Qy	1849	tcttgatgactggtcctcctgactctggttcctctatctactcgaagaagaatgctct	1908
Db	289	tgtttgacactggtcctcctgactctggttcctctatctactcgaagaagaatgctct	348
Qy	1909	gcaagaaccacaaagatcgatccgagaagaatcgatccacactccagacactgagcaaac	1968
Db	349	gcaaaaaccacaaagatcgatccgagaagaatcgatccacactccagacactgagcaaac	408
Qy	1969	cctgtctatacactcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag	2028
Db	409	ccctgtctatccatccatccagcagcagcagcagcagcagcagcagcagcagcagcagc	468
Qy	2029	cggttccaacattgtgacattcaacagcagcagcagcagcagcagcagcagcagcagcag	2088
Db	469	ctgtctccaacattgtgacattcaacagcagcagcagcagcagcagcagcagcagcagcag	528
Qy	2089	atgtctcaacctatgagaatcgatcgatcctgtgtatgacacacatgagcagcagcag	2148
Db	529	acgttttccatccatccatccagcagcagcagcagcagcagcagcagcagcagcagcag	588
Qy	2149	cagagctactgatacctgtgttgacaacatgagaaacacacacactagctgaactgaact	2208
Db	589	cagagctactgatacctgtgttgacaacatgagaaacacacacactagctgaactgaact	648
Qy	2209	tggttgggttacttgagcagcagcagcagcagcagcagcagcagcagcagcagcagcag	2268
Db	649	tggttgggttacttgagcagcagcagcagcagcagcagcagcagcagcagcagcagcag	708
Qy	2269	atccatctactacacagatctctcactggttccagctcagctgtgagcagtaactgac	2328



QY	2509	acctagctacatgectacatgltgtcctttgagatcaacgggaagatgtaaccactgaccc	2568
Db	954	ACGTGAGTACATGCGCCACTGTGTCTTGGATCAATGAGCAAAATGATCCACTGAC	1013
QY	2569	cctccgcctatccagagccagatcaagggttctgcaccggatgattccagatggaacc	2628
Db	1014	CCTCCGCTTATACCAAGCCAGACCAAGGCTTCTGTATCCAGTGGCTTCCAGAGTGAATTC	1073
QY	2629	attcccaagaatgagatcttggagagatgltcatcgtgtagatctactacacgctcttgaca	2688
Db	1074	ATTCCAGAAATGAGATCCGTGGGAGATGTTTTCATCCGAGATATATACACGCTTTTGACA	1133
QY	2689	gggccaacaacctcgctgggtctagctgaagaacatctga	2726
Db	1134	GGGCCAACCAACTCGTGGGGCTGGCCAAAGCATCTGGA	1171
RESULT	15		
LOCUS	A15633	1291 bp	mRNA linear PAT 18-FEB-1994
DEFINITION	preprochymosin.		
ACCESSION	A15633		
VERSION	A15633.1	GI:491951	
KEYWORDS			
SOURCE			
ORGANISM	Synthetic construct.		
REFERENCE	1 (bases 1 to 1291)		
AUTHORS	Carey,N.H., Doel,M.T., Harris,T.J.R., Lowe,P.A. and Emlage,J.S.		
TITLE	A process for the production of a polypeptide		
JOURNAL	Patent:EP 0068691-A 29 05-JAN-1983;		
FEATURES			
source	location/Qualifiers		
	1..1291		
gene	/organism="Synthetic construct"		
	/db_xref="taxon:32630"		
	26..1171		
	/gene="preprochymosin"		
CDS	26..1171		
	/gene="preprochymosin"		
	/codon_start=1		
	/transl_table=11		
	/protein_id="CAA01244.1"		
	/db_xref="GI:491952"		
	/translation="MRCLVLVLAFAFASQGAETIRPIYKGRSLRKALKEHGLDFPT		
	OKOQYGISKYSRGEVAVPLDTNVLDSQYKGIYLGPPDEFVLFPDSSDFWP		
	IYCKSNACKNHRPDRKRSSTFGNMGKLSIHVYSGMSOGLGIDVYVSNVIDLOO		
	VGLSDEPGEVFTVAEPDGLIGMAVPSLASYSIPIPENNMNRHLVADLPSSYMDRR		
	GOEMLTITGAINPSYSGSLHWVPVTVQYQNFYDSTVTSIGVYVACGGCQALIDRT		
	TSKLVPSSDLINTQATIGATONIGEDIDCDNLSIMPVFEINGRMPLTPSAVAT		
	SDDGPFSTSGFQSEHNSOKWILGDVIFREYYSVEDRANMLVGLAKAI"		
misc_difference	1291		
BASE COUNT	309 a 384 c 334 g 263 t	1 others	
ORIGIN			
Query Match	22.6%;	Score 894;	DB 6; Length 1291;
Best Local Similarity	87.5%;	Pred. No. 1.3e-131;	
Matches 978;	Conservative 0;	Mismatches 140;	Indels 0; Gaps 0;
QY	1609	tgtgttcggtactactagctgtgtgagatcaccgcgatctctctacaaagtaagcttc	1668
Db	54	TCTTGCTCTCTCCCAAGCGCTGAGATACCAGGATCCCTGTGACAAAGCAAGTCTC	113
QY	1669	tccgtgaaggcgtgaagaagacatggaactctgaagaagactcttgcagaacaacagatg	1728
Db	114	TGAGGAAGGCGCTGAAGGAGCATGGGCTTCTGAGAGACTCTCGACGAAMAACAGCAATG	173
QY	1729	gcataagcagaagtaactccggcttcogtgaagttgttagcgltgcacattacaactac	1788
Db	174	GCATCAGCAGCAAGTACTCTCGGCTTCGGGAGAGTGCCAGCGTCCCGGACCAACTACG	233

QY	1789	ttgtagtcaactcttgggaagaatccacctcggaaaccccgccctcaaggtltaaccgttc	1848
Db	234	TAGATAGTCACTACTTGGGAAGATCTACCTCGGAGACCCCGCCAGAGTTCCACGTGC	293
QY	1849	tctttgatactcgtgtcctctcgaacttcgtgttcctctatctactcgaaggaatgcct	1908
Db	294	TGTTTGGACACGTGGCTCTGACTTTCGGGTACCCCTCTATCTACATGCAAGACCAATGGCT	353
QY	1909	gcaagaaccaccagaagaatttcatccggagaagtcgfcacaccttcggaacttagcaaac	1968
Db	354	GCAAAAMCCACCACCGCTTGACCCGAGAAAGTCGTCCACTTCCAGACTTGGCCAAAC	413
QY	1969	ccctgtctatacactaagcgtgacagtgtaea tgcgaagaaactcttaggctatga taccgtca	2038
Db	414	CCCTGTCAATCACTACGCGGACAGGACGATGCAAGGCGATCCTAGAGATATGACACCGTTA	473
QY	2029	ctgtctccaacattgttgacaatcacaagacagtaagacttagcacccaagaacaggtg	2088
Db	474	CTGTCTCCAACTTGTGGAGATCCAGCAGACACTAGCCCTGAGCAACCCAAAGACCCGGG	533
QY	2089	atgtctcaacctatgcagaatctcgaatgcgtctccttgatcttgacataacca tgcgtcggt	2148
Db	534	ACGTTCTACCTATAGCCGAATTGACGCGGATCCTGGGGATGGCTACCTCCCTGCTCGCT	593
QY	2149	cagaagtactgataccctgtgttgacaaca tga tga accgaacactagtagtcaagact	2208
Db	594	CAGAGTACTGATACCCGTGTTGGACAACTGATGAACAGGCACTTGGTGCCCAAGAC	653
QY	2209	tgctctcgtgttlaa tga tga caagaatagccagaagaagcatgtcacaagctttagactatg	2268
Db	654	TGTTCTCGGTTTACATGAGACAGATTTGCTGAGGAGAGATGCTCAACGCTGGGGGCCATCA	713
QY	2269	atccatccctactacaagga tctcctcctaacgtgggttcacgaactctgacagagactggc	2328
Db	714	ACCCGTCTCTACTACACAGGGTCCCTGCACTGGGTGCCCTGACATGCAAGCATCTGAC	773
QY	2329	aattcaactgtagacagtgtcaccaatcaagcaggtgtgtgtgtgtgtgcatgtgaagttgtagtc	2388
Db	774	AGTTCACTAGTGGACAGTGTCAACCATCAGCGGTGTGTAATGAGCTGTGAGGGTGGCTGTC	833
QY	2389	aagctatcttggatacccggttaag tccaagctgtgttgagacctagcagcgacatctccaaca	2448
Db	834	AGGCGATCTTGGACACGGGACCTCCAAAGCTGGTCCGGGCCACGACGCAATCCTCAACA	893
QY	2449	ttcgaagaagctattggagcgacacacaagaacagtagtgtaagttacatagatgtgcaga	2508
Db	894	TCCACGACGAGCATTTGGAGCCACACAGAACCATGACGTGAAGTTTGACATCGACTCCGACA	953
QY	2509	accttagctatactgctcactagttgtcctttgagatacaagcgacaagatgtlaaccactgaacc	2568
Db	954	ACCTGAGCTACATGCCCACTGTTGGTCTTTGAGATCAATGGCAAAATGTACCACATGACCC	1013
QY	2559	ccctcgcgctataccagcgacagatccaaggtgttcgcacacagtgatccaagagtgaacc	2628
Db	1014	CCCTCGGCTATACCAAGCAAGACCAAGCGCTTCTGTACAGTGGCTTCCAGATGTAAATTC	1073
QY	2629	attcccaagaatgatattcttggagatgtgttcaatcgtttagtaactacaagcgtcttttga	2688
Db	1074	ATTTCCAGAAATGATCTCTGGGGGATGTTTTCATCCGAAGATATTACACGCTTTTGACA	1133
QY	2689	gggcacaacaactcgttgggtcgtagtcaagaacatctga	2726
Db	1134	GGGCAACAACCTCTGGGGCTGCGCCAAAGCCATCTGCA	1171

Search completed: July 31, 2002, 18:19:33  
Job time: 20124 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 21:11:57 ; Search time 790.32 Seconds  
(without alignments)  
8596.303 Million cell updates/sec

Title: US-09-643-755B-3

Perfect score: 3957

Sequence: 1 ctgcaggaattcatgtact.....acctaccactaagtacc 3957

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

N\_Geneseq\_032802:\*

- 1: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*
- 2: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*
- 3: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*
- 4: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*
- 5: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*
- 6: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*
- 7: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*
- 8: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*
- 9: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*
- 10: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*
- 11: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*
- 12: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*
- 13: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*
- 14: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*
- 15: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*
- 16: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*
- 17: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*
- 18: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*
- 19: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*
- 20: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*
- 21: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*
- 22: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*
- 23: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*
- 24: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	3957	100.0	22	AAS00570
2	1545.2	39.0	20	AAZ10376
3	1539	38.9	20	AAZ10392
4	1226.8	31.0	20	AAZ10392
5	1173	29.6	22	AAS00569
6	1149.8	29.1	22	AAZ17528
7	898.8	22.7	10	AAZ1157
8	897.4	22.7	16	AAZ03006
9	897.2	22.7	5	AAZ40295

10	897.2	22.7	1460	5	AAZ40180	Sequence of recomb
11	896.4	22.7	1210	12	AAZ14051	Remain gene. Synt
12	895.8	22.6	2733	20	AAZ06463	2.7 Kbp HindIII fr
13	895.6	22.6	1278	5	AAZ40055	Sequence of prochy
14	895.6	22.6	1460	3	AAZ20043	Pre-prorenin-A ge
15	895.6	22.6	2727	10	AAZ11188	BamHI/SalI insert
16	894	22.6	1290	4	AAZ30209	Sequence of prepro
17	890.8	22.5	1314	4	AAZ30049	CDNA sequence cor
18	889.8	22.5	1289	4	AAZ30022	Sequence of veal c
19	886.8	22.4	1098	11	AAZ04683	Sequence encoding
20	885.2	22.4	1098	4	AAZ30063	Sequence of proten
21	875.4	22.1	2982	10	AAZ1185	BamHI insert from
22	866.6	21.9	1175	13	AAZ20949	Prochymosin (proe
23	704.6	17.8	1143	5	AAZ40214	Gene encoding (pre
24	686.8	17.4	1104	9	AAZ80001	Optimised prochymo
25	534	13.5	1140	24	AAZ97159	Human aspartyl pro
26	382.4	9.7	637	14	AAZ49459	Prochymosin gene 5
27	359.2	9.1	1393	22	AAZ57421	Human stomach cell
28	359.2	9.1	1409	22	AAZ44608	Human full-length
29	315.4	8.0	1017	23	AAZ79579	DNA encoding novel
30	291	7.4	1360	20	AAZ20193	Cat pregnancy asso
31	248.4	6.3	1357	22	AAZ57422	Human stomach cell
32	248.4	6.3	1366	22	AAZ09486	Human gastricsin c
33	248.4	6.3	1418	20	AAZ77523	Human ovarian tumo
34	247.6	6.0	1285	20	AAZ20179	Bovine pregnancy a
35	236.2	5.9	1258	20	AAZ20181	Bovine pregnancy a
36	234.6	5.9	1258	20	AAZ20162	Bovine pregnancy a
37	231.6	5.9	1130	20	AAZ20182	Bovine pregnancy a
38	230.4	5.8	1173	20	AAZ20183	Bovine pregnancy a
39	226.8	5.7	1173	24	AAZ97161	Human aspartyl pro
40	221	5.6	1358	20	AAZ39788	Gastric cancer ass
41	203.8	5.2	1180	23	ABL21921	Drosophila melanog
42	203.8	5.2	3180	23	ABL21920	DNA fragment of pa
43	203	5.1	828	11	AAZ03224	DNA fragment of pa
44	199.8	5.0	846	11	AAZ03223	DNA fragment of pa
45	197.2	5.0	1168	20	AAZ20171	Bovine pregnancy a

#### ALIGNMENTS

RESULT	ID	Score	Query Match Length	ID	Description
1	AAS00570	3957	100.0	22	AAS00570
2	1545.2	39.0	20	AAZ10376	Bovine phaseolin p
3	1539	38.9	20	AAZ10392	Nucleotide sequenc
4	1226.8	31.0	20	AAZ10392	Phaseolin vulgaris
5	1173	29.6	22	AAS00569	Nucleotide sequenc
6	1149.8	29.1	22	AAZ17528	Bovine pre-pro-chy
7	898.8	22.7	10	AAZ1157	Kunitz soybean try
8	897.4	22.7	16	AAZ03006	Cloned sequence of
9	897.2	22.7	5	AAZ40295	Chymosin open read

PD 01-MAR-2001.  
XX  
PF 23-AUG-2000; 2000WO-CA00975.  
XX  
PR 23-AUG-1999; 99US-0378696.  
XX  
FA (SEMB-) SEMBIOSYS GENETICS INC.  
XX  
PI Van Rooijen G, Keon RG, Boothe J, Shen Y;  
XX  
DR WPI: 2001-226621/23.  
DR P-PSDB; AAU00536.  
XX  
PT Producing chymosin in seeds of plants such as rice, flax, rape seed, by  
PT transforming plant cell with a nucleic acid encoding chymosin operably  
PT linked to transcription regulator and terminator sequences -  
XX  
PS Example 1; Fig 2; 56pp; English.  
XX  
CC The sequence represents a chimeric polynucleotide comprising a  
CC pre-pro-chymosin, a phaseolin promoter and a phaseolin terminator.  
CC Chymosin can be produced in a plant seed through introduction of a  
CC chimeric nucleic acid molecule, comprising a nucleic acid sequence  
CC encoding a chymosin polypeptide operatively linked to transcription  
CC regulator and terminator sequences, into a plant cell. The sequences are  
CC useful for producing plant seeds, in particular seeds of soybean, rape  
CC seed, sunflower, cotton, corn, tobacco, alfalfa, wheat, barley, oats,  
CC sorghum, Arabidopsis thaliana, potato, flax/linseed, safflower, oil palm,  
CC groundnut, Brazil nut, coconut, castor, coriander, squash, jojoba and  
CC rice.  
XX  
SQ Sequence 3957 BP; 1263 A; 790 C; 609 G; 1295 T; 0 other;

Query Match 100.0%; Score 3957; DB 22; Length 3957;  
Best Local Similarity 100.0%; Pred. NO. 0;  
Matches 3957; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctgcaggaattcatgtactccagatcatcattatagtagaagtttgctctcgcgcg 60  
DB 1 ctgcaggaattcatgtactccagatcatcattatagtagaagtttgctctcgcgcg 60  
QY 61 tggtttttaacctattaaagggtttccacctaaattctggtatcttctact 120  
DB 61 tggtttttaacctattaaagggtttccacctaaattctggtatcttctact 120  
QY 121 ttactgttacttaattctcataatcttggttgaaatttccagcttcgcgcacga 180  
DB 121 ttactgttacttaattctcataatcttggttgaaatttccagcttcgcgcacga 180  
QY 181 tatccctcaaatattatattgtttaaacatttcaaacgcgacataaatttatgaagtc 240  
DB 181 tatccctcaaatattatattgtttaaacatttcaaacgcgacataaatttatgaagtc 240  
QY 241 ccgtctatcttaagtgtactacaatttcaatgtgaaatatataattacttaattt 300  
DB 241 ccgtctatcttaagtgtactacaatttcaatgtgaaatatataattacttaattt 300  
QY 301 agcggttggagaagaacataaagatttatcttatcttccatcttaaatgtttatata 360  
DB 301 agcggttggagaagaacataaagatttatcttatcttccatcttaaatgtttatata 360  
QY 361 caataaacaacattcttacccttaagaagattcccatcttataatttaanaatalatc 420  
DB 361 caataaacaacattcttacccttaagaagattcccatcttataatttaanaatalatc 420  
QY 421 ttatacaaatattttcaaccacgtaaatctcataaataaagttggttcaaaagtaata 480  
DB 421 ttatacaaatattttcaaccacgtaaatctcataaataaagttggttcaaaagtaata 480  
QY 481 aatttaactccaaattttttatctcgaactgatacttaagaacaccccgtagacaact 540  
DB 481 aatttaactccaaattttttatctcgaactgatacttaagaacaccccgtagacaact 540

QY 541 agccattttttcttltgataaaaaatccaatatcatctgtattttttatatacaatga 600  
DB 541 agccattttttcttltgataaaaaatccaatatcatctgtattttttatatacaatga 600  
QY 601 aaattcaccaacaacatactatgtgtatcttctgaagcaagtcagtatgcaaatctc 660  
DB 601 aaattcaccaacaacatactatgtgtatcttctgaagcaagtcagtatgcaaatctc 660  
QY 661 ataattccatttgcactacggaagtaacggaatcgtcttaccatgtagagacacat 720  
DB 661 ataattccatttgcactacggaagtaacggaatcgtcttaccatgtagagacacat 720  
QY 721 ctctcaagtaattttaataagtagtactatataattcaagatttcataataactcaaa 780  
DB 721 ctctcaagtaattttaataagtagtactatataattcaagatttcataataactcaaa 780  
QY 781 tattctctcaaaaaaataatagataataataatattacttttttaatttaagttt 840  
DB 781 tattctctcaaaaaaataatagataataataatattacttttttaatttaagttt 840  
QY 841 aatttgaatttggactattgtatttattctactatgtttaattgtttatagat 900  
DB 841 aatttgaatttggactattgtatttattctactatgtttaattgtttatagat 900  
QY 901 agtttaagtaataataagtaatgtagtagagtgtagagtgtagagtgtagagtgtagag 960  
DB 901 agtttaagtaataataagtaatgtagtagagtgtagagtgtagagtgtagagtgtagag 960  
QY 961 tataacatttaagtgtagactaatttccatataatttcttattgtctttacatttctgt 1020  
DB 961 tataacatttaagtgtagactaatttccatataatttcttattgtctttacatttctgt 1020  
QY 1021 atgtatgcgtaactagaattacagtggttgcacatggaactcgtgtcttggcttc 1080  
DB 1021 atgtatgcgtaactagaattacagtggttgcacatggaactcgtgtcttggcttc 1080  
QY 1081 atgcattggtcttgcgcaagaaagacaaagaaagaaagaaagaaagaaagaaagaa 1140  
DB 1081 atgcattggtcttgcgcaagaaagacaaagaaagaaagaaagaaagaaagaaagaa 1140  
QY 1141 caaaacgcacatcaccaac 1200  
DB 1141 caaaacgcacatcaccaac 1200  
QY 1201 atgtatgcttaaatgccatgcaagcaaacgtgttlaacatgcaactttaaattgctcac 1260  
DB 1201 atgtatgcttaaatgccatgcaagcaaacgtgttlaacatgcaactttaaattgctcac 1260  
QY 1261 ccactcacaaccacaacaacaacacacacacacacacacacacacacacacacacac 1320  
DB 1261 ccactcacaaccacaacaacaacacacacacacacacacacacacacacacacacac 1320  
QY 1321 atatatattcttcttcgcgacccctcaatttcttcaattcaacaacagtcacacacac 1380  
DB 1321 atatatattcttcttcgcgacccctcaatttcttcaattcaacaacagtcacacacac 1380  
QY 1381 tgggtgtacatcccatgccccaaatctccatgcatggttccaaaccttctcttatataa 1440  
DB 1381 tgggtgtacatcccatgccccaaatctccatgcatggttccaaaccttctcttatataa 1440  
QY 1441 tacctataaatacctctaatatcactcaacttcttcaatccatccatccacagtagta 1500  
DB 1441 tacctataaatacctctaatatcactcaacttcttcaatccatccatccacagtagta 1500  
QY 1501 ctactctactataataaccccccaacacacacacacacacacacacacacacacacacac 1560  
DB 1501 ctactctactataataaccccccaacacacacacacacacacacacacacacacacacac 1560  
QY 1561 tcccttaagcttcccttctcagcttcccttcttcttcttcttcttcttcttcttcttct 1620  
DB 1561 tcccttaagcttcccttctcagcttcccttcttcttcttcttcttcttcttcttcttct 1620



QY 1621 ctcaacgctgctgagatcacccgcgcatctctctacaagaagtcaagtcctccgtaaagcgc 1680  
|||||  
Db 1621 ctcaacgctgctgagatcacccgcgcatctctctctacaagaagtcaagtcctccgtaaagcgc 1680  
QY 1661 tgaaggaacatgagacttcttaagaactctctgcagaaacaacagatgacatgacagaca 1740  
|||||  
Db 1661 tgaaggaacatgagacttcttaagaactctctgcagaaacaacagatgacatgacagaca 1740  
QY 1741 agtactccggcttcgctgtaagtctgctagcgtgcacacttacaacacttgaatgaat 1800  
|||||  
Db 1741 agtactccggcttcgctgtaagtctgctagcgtgcacacttacaacacttgaatgaat 1800  
QY 1801 acttgggaagaatcatctaccccggaaccccgctccaagaatgtaacgcttctcttgaatg 1860  
|||||  
Db 1801 acttgggaagaatcatctaccccggaaccccgctccaagaatgtaacgcttctcttgaatg 1860  
QY 1861 gtctccttgactcttggtgtccctctatctactgcaagaagcaatgctgcagaagaccac 1920  
|||||  
Db 1861 gtctccttgactcttggtgtccctctatctactgcaagaagcaatgctgcagaagaccac 1920  
QY 1921 aaagatctgcgacggaagaagtcgtccacacttccagaacttgaagcaacccctgtctatac 1980  
|||||  
Db 1921 aaagatctgcgacggaagaagtcgtccacacttccagaacttgaagcaacccctgtctatac 1980  
QY 1981 actacgctgacagtgagcatgcaagaatcttaagctatgataccgctacgtctccaaca 2040  
|||||  
Db 1981 actacgctgacagtgagcatgcaagaatcttaagctatgataccgctacgtctccaaca 2040  
QY 2041 ttgtgagacattcaacagacagtaaggaacttagcaacccaagaacagtgatgtctacact 2100  
|||||  
Db 2041 ttgtgagacattcaacagacagtaaggaacttagcaacccaagaacagtgatgtctacact 2100  
QY 2101 atgcagaacttgatgagacacgtctgtgatgcataccacacgtctgcgcgaagatctga 2160  
|||||  
Db 2101 atgcagaacttgatgagacacgtctgtgatgcataccacacgtctgcgcgaagatctga 2160  
QY 2161 taactgtgttgtaacacatga tgaaccgcacacactagtagccccaagaactgtctcgtgtt 2220  
|||||  
Db 2161 taactgtgttgtaacacatga tgaaccgcacacactagtagccccaagaactgtctcgtgtt 2220  
QY 2221 acatgagacaaggaatgycgacagagagcatgctcaacgcttgagctatgatccatccact 2280  
|||||  
Db 2221 acatgagacaaggaatgycgacagagagcatgctcaacgcttgagctatgatccatccact 2280  
QY 2281 acacaagatctcttcacactggttccagtcactgtgcaagcaagtaactggaattcatctg 2340  
|||||  
Db 2281 acacaagatctcttcacactggttccagtcactgtgcaagcaagtaactggaattcatctg 2340  
QY 2341 acagtgtaaccatcacgctgtgtgtgtgtgtgcatgtgaagtgagatgcaactcatctg 2400  
|||||  
Db 2341 acagtgtaaccatcacgctgtgtgtgtgtgtgcatgtgaagtgagatgcaactcatctg 2400  
QY 2401 ataccgtaagtcacaagctgtgctggaacctagcaagcgaacatctccaacatcaagaagcta 2460  
|||||  
Db 2401 ataccgtaagtcacaagctgtgctggaacctagcaagcgaacatctccaacatcaagaagcta 2460  
QY 2461 ttgagagccacagacaacacagtaagctgtgattgacatagattgcgacaacttgaactaca 2520  
|||||  
Db 2461 ttgagagccacagacaacacagtaagctgtgattgacatagattgcgacaacttgaactaca 2520  
QY 2521 tgcctacagttgtctttgagatcaacgycgaagatgtaaccactgacccctccgcctta 2580  
|||||  
Db 2521 tgcctacagttgtctttgagatcaacgycgaagatgtaaccactgacccctccgcctta 2580  
QY 2581 ccagccagatccaaggtgtctgcacacagtgattccagagtgaagacattccaccagaat 2640  
|||||  
Db 2581 ccagccagatccaaggtgtctgcacacagtgattccagagtgaagacattccaccagaat 2640  
QY 2641 ggaacttggagagatgtgtcaatctgtgtagtactacaagcgtctttgacagggccaacaac 2700  
|||||  
Db 2641 ggaacttggagagatgtgtcaatctgtgtagtactacaagcgtctttgacagggccaacaac 2700  
QY 2701 tcgttggctagctaaagaacatctgaagcttaataagtaagtaactaaatgcatgtagt 2760  
|||||

Db 2701 tcgttggctagctaaagaacatctgaagcttaataagtaagtaactaaatgcatgtagt 2760  
|||||  
QY 2761 gtaagaagctcaatgagagacaatggaatatgtatccgaccacgtacaacatataataacgta 2820  
|||||  
Db 2761 gtaagaagctcaatgagagacaatggaatatgtatccgaccacgtacaacatataataacgta 2820  
QY 2821 gttccatccacttctctatgaataaacaagagtgatagatatataaacactatc 2880  
|||||  
Db 2821 gttccatccacttctctatgaataaacaagagtgatagatatataaacactatc 2880  
QY 2881 tatgacacttattgttctatgataaattcccttattatataataatcatctgaatctg 2940  
|||||  
Db 2881 tatgacacttattgttctatgataaattcccttattatataataatcatctgaatctg 2940  
QY 2941 acggtctatggaatgcttccaatagtaacaaaacaacatgtaactataagacttctaa 3000  
|||||  
Db 2941 acggtctatggaatgcttccaatagtaacaaaacaacatgtaactataagacttctaa 3000  
QY 3001 caattctaaccttagcatgtggaacgagacataaagtgttaagaagacataacaattataa 3060  
|||||  
Db 3001 caattctaaccttagcatgtggaacgagacataaagtgttaagaagacataacaattataa 3060  
QY 3061 tggagaaggttgtctccattatataatataataatccacttattgatatataagga 3120  
|||||  
Db 3061 tggagaaggttgtctccattatataatataataatccacttattgatatataagga 3120  
QY 3121 tgttaagagacataacaacatataaagaagagatgttatccattatataatataac 3180  
|||||  
Db 3121 tgttaagagacataacaacatataaagaagagatgttatccattatataatataac 3180  
QY 3181 taaccattatataatataatataatccattatataatgttataaagtttgaacgt 3240  
|||||  
Db 3181 taaccattatataatataatataatccattatataatgttataaagtttgaacgt 3240  
QY 3241 atatttccaatatatttagttgatatgtatgaatgaagggttactatttgaactcttactc 3300  
|||||  
Db 3241 atatttccaatatatttagttgatatgtatgaatgaagggttactatttgaactcttactc 3300  
QY 3301 tgtataaaggttggatcaactcccttaagtggtgtccttaatttattgtcttccacagat 3360  
|||||  
Db 3301 tgtataaaggttggatcaactcccttaagtggtgtccttaatttattgtcttccacagat 3360  
QY 3361 aaaaaaaaatataagttgtgttgaataaatattgaaggaatttaataataataata 3420  
|||||  
Db 3361 aaaaaaaaatataagttgtgttgaataaatattgaaggaatttaataataataataata 3420  
QY 3421 ataaataacataataatataatgataataatattatataataataacattatctataaanaa 3480  
|||||  
Db 3421 ataaataacataataatataatgataataatattatataataataacattatctataaanaa 3480  
QY 3481 agtaaatatgtcaataaactataacacgctttagccttgycgagacctccaattatc 3540  
|||||  
Db 3481 agtaaatatgtcaataaactataacacgctttagccttgycgagacctccaattatc 3540  
QY 3541 taaagagagtaaaacataattggaactttgttatttaacaacattatatttaaacact 3600  
|||||  
Db 3541 taaagagagtaaaacataattggaactttgttatttaacaacattatatttaaacact 3600  
QY 3601 atgaattttttttttatccgcaaggaataataaatttaagagaggaacatgct 3660  
|||||  
Db 3601 atgaattttttttttatccgcaaggaataataaatttaagagaggaacatgct 3660  
QY 3661 gtcccaatccctatatacaacacactccacaggaagtgatgcggygcacacaaaaaac 3720  
|||||  
Db 3661 gtcccaatccctatatacaacacactccacaggaagtgatgcggygcacacaaaaaac 3720  
QY 3721 aggcagaggaatttttaatttgggt 3780  
|||||  
Db 3721 aggcagaggaatttttaatttgggt 3780  
QY 3781 caatacacataacccctttagcaagtagagcaatgtgtacccgtgtgcttagcttctta 3840  
|||||

Db 3781 cactacataaaccttttagcagtagagcaatggtgaccggtgcttagcttcttla 3840  
Qy 3841 tttaattttttatcagaagaataataataataatgagacactcagaagtcttc 3900  
|||||  
Db 3841 tttaattttttatcagaagaataataataataatgagacactcagaagtcttc 3900  
Qy 3901 aacctatatacaaaaccccaaaacaaagtcttcctagcaccctaccaactaagtacc 3957  
|||||  
Db 3901 aacctatatacaaaaccccaaaacaaagtcttcctagcaccctaccaactaagtacc 3957  
RESULT 2  
AAZ10376  
ID AAZ10376 standard; DNA; 1558 BP.  
XX  
AC AAZ10376;  
XX  
DT 15-NOV-1999 (first entry)  
XX  
DE Nucleotide sequence of the bean phaseolin promoter.  
XX  
KW Fao gene: plant metabolism; fatty acid oxidation enzyme;  
KW polyhydroxyalkanoate; oil composition; seed production; plant biomass;  
KW transgenic plant; promoter; bean; ss.  
XX  
OS Phaseolus sp.  
XX  
PN W09945122-A1.  
XX  
PD 10-SEP-1999.  
XX  
PF 05-MAR-1999; 99WO-US04999.  
XX  
PR 06-MAR-1998; 98US-0077107.  
XX  
PA (META-) METABOLIX INC.  
XX  
PI Boynton L, Huismann GW, Moloney M, Patterson N, Peoples OP;  
PI Snell K;  
XX  
WP: 1999-540850/45.  
XX  
PT Modifying fatty acid metabolism in plants, useful for increasing  
PT biomass and producing specific polymers in seeds  
XX  
PS Example 3: Page 63; 79pp; English.  
XX  
CC The present sequence represents the bean phaseolin promoter, which  
CC is used to construct plasmids for the expression of the P. putida  
CC facAB gene. The gene encodes an enzyme that may be used in the  
CC method of the invention. The specification describes a method for  
CC manipulating the metabolism of a plant, and comprises expressing a  
CC heterologous gene encoding fatty acid oxidation enzymes in the  
CC cytosol or plastids other than the peroxisomes, glyoxisomes or  
CC mitochondria of the plant. The method may be used to enhance the  
CC biological production of polyhydroxyalkanoates or novel oil compositions  
CC in a transgenic plant. Plants which may be used to produce these  
CC compounds in this way include Brassicas, maize, soybean, cottonseed,  
CC sunflower, palm, coconut, safflower, peanut, mustard, flax, tobacco and  
CC alfalfa. The method may also be used to prevent or suppress seed  
CC production and therefore increase the production of biomass (leaves,  
CC stems, stalks) by plants.  
CC  
SQ Sequence 1558 BP; 516 A; 310 C; 171 G; 561 T; 0 other;

Qy 66 tttaaccttaatttaagggttttccacctaataaattcgtatcatctcacttact 125  
|||||  
Db 62 tttaaccttaatttaagggttttccacctaataaattcgtatcatctcacttact 121  
Qy 126 tgttaacttaatttcctcaataatcttggttgtaaatattcaagcttcgcgacagatc 185  
|||||  
Db 122 tgttaacttaatttcctcaataatcttggttgtaaatattcaagcttcgcgacagatc 181  
Qy 186 ctacaatattatatttgatttaaacatttccaacgcgataaatttatagaagccgcg 245  
|||||  
Db 182 ctacaatattatatttgatttaaacatttccaacgcgataaatttatagaagccgcg 241  
Qy 246 tatcttaatgtagtctaacatttccatbatgaaatataataattacttaattgaagc 305  
|||||  
Db 242 tatcttaatgtagtctaacatttccatbatgaaatataataattacttaattgaagc 301  
Qy 306 tggtagaagacataaagaattatcttattcttcttccataaagtgttaataataata 365  
|||||  
Db 302 tggtagaagacataaagaattatcttattcttcttccataaagtgttaataataata 361  
Qy 366 taacaataattcttaaccttaagaagatttccatttataatttaaaaaatataatc 425  
|||||  
Db 362 taacaataattcttaaccttaagaagatttccatttataatttaaaaaatataatc 421  
Qy 426 aaataatttcaacacgtaaatctcabaataaagtgttccaagaataaattt 485  
|||||  
Db 422 aaataatttcaacacgtaaatctcabaataaagtgttccaagaataaattt 481  
Qy 486 aacctcaaatttttttatctgactgacttaagaagacacccagtgacaaactagcca 545  
|||||  
Db 482 aacctcaaatttttttatctgactgacttaagaagacacccagtgacaaactagcca 541  
Qy 546 ttttttcttggataaaaaaatccaattatcatgttattttttatatacaatgaaatt 605  
|||||  
Db 542 ttttttcttggataaaaaaatccaattatcatgttattttttatatacaatgaaatt 601  
Qy 606 tcaaccaacaatcatgttggatctctgaagcaagtcagtattatgcaaaattcctaat 665  
|||||  
Db 602 tcaaccaacaatcatgttggatctctgaagcaagtcagtattatgcaaaattcctaat 661  
Qy 666 tccatttgacactacggaagtactggaagctcgttttcaatgagcgagacactctct 725  
|||||  
Db 662 tccatttgacactacggaagtactggaagctcgttttcaatgagcgagacactctct 721  
Qy 726 aaagtaatttaataatagttactatataatccaagattcatalatcaatcaatata 785  
|||||  
Db 722 aaagtaatttaataatagttactatataatccaagattcatalatcaatcaatata 781  
Qy 786 ctcttaaaaaatataatagataataataaataattactttttaaatttaagttaattg 845  
|||||  
Db 782 ctcttaaaaaatataatagataataataaataattactttttaaatttaagttaattg 841  
Qy 846 ttgaatttgtagacttattattattattatctctatcattggttaattggtttatagatg 905  
|||||  
Db 842 ttgaatttgtagacttattattattattatctctatcattggttaattggtttatagatg 901  
Qy 906 aaagtaaatataaagtaagtagtagagtgtagagtgtagccttaaacataaactataa 965  
|||||  
Db 902 aaagtaaatataaagtaagtagtagagtgtagagtgtagccttaaacataaactataa 961  
Qy 966 cattatgtagactaatttccatatattcttattcttacttgccttcttgtagtata 1025  
|||||  
Db 962 cattatgtagactaatttccatatattcttattcttacttgccttcttgtagtata 1021  
Qy 1026 agtcgtaactagaattacagtggttgccatgagactctggtgcttttgtagtata 1085  
|||||  
Db 1022 agtcgtaactagaattacagtggttgccatgagactctggtgcttttgtagtata 1081  
Qy 1086 tgggtcttgccgcaaaaaagacaagaacaagaaaaaagacaagaagagagaca 1145  
|||||  
Db 1082 tgggtcttgccgcaaaaaagacaagaacaagaaaaaagacaagaagagagaca 1141

QY 1146 cgcatacacacaaccaactcaatagtagtactggtgatacgaatgcgcgcgtccatgta 1205  
|||||  
Db 1142 cgcatacacacaaccaactcaatagtagtactggtgatacgaatgcgcgcgtccatgta 1201  
QY 1206 tgtctaataatgcgcgaagaagaacagtgcttaacatgcaacttaaatggtcacccatc 1265  
|||||  
Db 1202 tgcctaataatgcgcgaagaagaacagtgcttaacatgcaacttaaatggtcacccatc 1261  
QY 1266 tcaacccccacaacaacacatgctctttcttctatcatcacccaacacacgtatata 1325  
|||||  
Db 1262 tcaacccccacaacaacacatgctctttcttctatcatcacccaacacacgtatata 1321  
QY 1326 ttcatctcttcgcgcacactcaattcttctcaacttcaacacagtcacactgcatagct 1385  
|||||  
Db 1322 ttcatctcttcgcgcacactcaattcttctcaacttcaacacagtcacactgcatagct 1381  
QY 1386 gtatcccatgcccacaatcttcacatgcatggtccaaacacctctctctatataatacct 1445  
|||||  
Db 1382 gtatcccatgcccacaatcttcacatgcatggtccaaacacctctctctatataatacct 1441  
QY 1446 ataataacctataatcaactcttcttcatcatcatcatcatccagagtaactact 1505  
|||||  
Db 1442 ataataacctataatcaactcttcttcatcatcatcatcatccagagtaactact 1501  
QY 1506 ctactactataatacccaaccaactcatatactcaatactactactat 1555  
|||||  
Db 1502 ctactactataatacccaaccaactcatatactcaatactactactat 1551

RESULT 3  
AAx90961 standard; DNA; 1547 BP.  
XX AAX90961;  
XX 17-JAN-2000 (first entry)  
XX Phaseolus vulgaris beta phaseolin gene 5' regulatory sequence.  
XX  
XX Palmitate-CoA delta-9 desaturase; PCA; beta phaseolin gene; fatty acid;  
XX palmitic acid; desaturase; modification; saturation; oil; transgenic plant;  
XX 5-regulatory element; plasmid phagN184-2; tobacco; transformed plant; ss.  
XX  
XX Phaseolus vulgaris.  
XX  
XX Key Location/Qualifiers  
XX old\_sequence /tag= a  
XX /note= "Corresponds to base 1113 of plasmid phagN184-2  
XX and facilitates subsequent cloning"  
XX  
XX W09950430-A2.  
XX 07-OCT-1999.  
XX  
XX 29-MAR-1999; 99WO-US06765.  
XX  
XX 30-MAR-1998; 98US-0079840.  
XX  
XX (DOWC ) DOW AGRSCIENCES LLC.  
XX  
XX Folkerts O, Merlo DJ;  
XX  
XX WPI; 1999-610860/52.  
XX  
XX  
XX New isolated palmitoyl-CoA delta-9 desaturase nucleic acids, used to  
XX produce transgenic plants with altered levels of fatty acids  
XX  
XX Example 2; Page 68-69; 78pp; English.  
XX  
XX The present sequence is the 5' regulatory sequence of the beta phaseolin  
XX gene from Phaseolus vulgaris. It is used in the construction of the  
XX plasmid phagN184-2. This regulatory sequence is suitable for high level

CC and seed specific expression in tobacco. This plasmid can be used for  
CC producing transgenic plants having altered saturated oil profiles and  
CC decreased fatty acid levels in transformed plants.  
XX  
XX Sequence 1547 BP; 513 A; 306 C; 170 G; 558 T; 0 other.  
Query Match 38.9%; Score 1539; DB 20; Length 1547;  
Best Local Similarity 99.7%; Pred. No. 1.4e-208;  
Matches 1542; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
QY 6 ggaattctgtagtaccagatcatatagtagaagtttggctctcgcgcggttggt 65  
| | | | |  
Db 1 gaattccattgacctccagatcatatagtagaagtttggctctcgcgcggttggt 60  
QY 66 tttaacctatttaaaagggtttccacactaaattctgtagatcttcaacttact 125  
| | | | |  
Db 61 tttaacctatttaaaagggtttccacactaaattctgtagatcttcaacttact 120  
QY 126 tgtacttaatttctcatatccttggttggaattatcaacgcttcgcacagatcc 185  
| | | | |  
Db 121 tgtacttaatttctcatatccttggttggaattatcaacgcttcgcacagatcc 180  
QY 186 ctacaatttatattgtttaaacaatttcaaacgcgaataaattttaagagtcgcgc 245  
| | | | |  
Db 181 ctacaatttatattgtttaaacaatttcaaacgcgaataaattttaagagtcgcgc 240  
QY 246 tatctttaatgtagtctaacatttcatatagaaatataataatttacttaattagcgt 305  
| | | | |  
Db 241 tatctttaatgtagtctaacatttcatatagaaatataataatttacttaattagcgt 300  
QY 306 tggtagaagaataagaattatcttcttctctcatataaagtttaataatcata 365  
| | | | |  
Db 301 tggtagaagaataagaattatcttcttctctcatataaagtttaataatcata 360  
QY 366 taaacaattcttacccttaagaagaattcccatatttataatttaaaatatatttc 425  
| | | | |  
Db 361 taaacaattcttacccttaagaagaattcccatatttataatttaaaatatatttc 420  
QY 426 aaatatcttcaacacagtaaatctcaataataaagtgttttcaaaagtaataatt 485  
| | | | |  
Db 421 aaatatcttcaacacagtaaatctcaataataaagtgttttcaaaagtaataatt 480  
QY 486 aactcataattttttatcgactgactttaagaacacacccagtgacaacagca 545  
| | | | |  
Db 481 aactcataattttttatcgactgactttaagaacacacccagtgacaacagca 540  
QY 546 ttttttcttgaataaaaaaaccaattatcatgtattttttatatacagaatgaatt 605  
| | | | |  
Db 541 ttttttcttgaataaaaaaaccaattatcatgtattttttatatacagaatgaatt 600  
QY 606 tcaaccaacaatcattgtggtgatttctgaagcaagtcagttaagcaaaattcataat 665  
| | | | |  
Db 601 tcaaccaacaatcattgtggtgatttctgaagcaagtcagttaagcaaaattcataat 660  
QY 666 tcccatctgacactaaggaaatgaatgaagcttgcatttcatgycggaacacactct 725  
| | | | |  
Db 661 tcccatctgacactaaggaaatgaatgaagcttgcatttcatgycggaacacactct 720  
QY 726 aaagtaattttaataatagttactatatacagaattcatatatacaactcaactata 785  
| | | | |  
Db 721 aaagtaattttaataatagttactatatacagaattcatatatacaactcaactata 780  
QY 786 ctcttaaaaatataatgataataataaataattacttttaatttaagttaattg 845  
| | | | |  
Db 781 ctcttaaaaatataatgataataataaataattacttttaatttaagttaattg 840  
QY 846 ttgaatttgacattatattatattatctactactaagtttaattggttttagagtt 905  
| | | | |  
Db 841 ttgaatttgacattatattatattatctactactaagtttaattggttttagagtt 900  
QY 906 aaagtaataataatgataatgtagaggttagaggttacccttaaacataataataa 965  
| | | | |

Dg	901	aaagtaataataagtaatgatagtatgagtgtaagtgtaagtgtaacccaataacataactaa	960
Qy	966	catttatggtgagcctaattttcataabatttctaatttgcttctaacttcttcttgta	1025
Db	961	gattatggtgagcctaattttcataabatttctaatttgcttctaacttcttcttgta	1020
Qy	1026	agtcgcgaactaataattacaaagtgggtttgcgcagagcctgttgcttttggtcatgca	1085
Db	1021	agtcgcgaactaataattacaaagtgggttttgtaagagcaccctgttgcttttggtcatgca	1080
Qy	1086	tgggtttcttgcaagaagaaagacaaagaaacaaagaaaaagacaaaagagagacaaa	1145
Db	1081	tgggtttcttgcaagaagaaagacaaagaaacaaagaaaaagacaaaagagagacaaa	1140
Qy	1146	cgcgaatcaccaaccacactccaattagtctagctgcttgataagatcgcggtccatgta	1205
Db	1141	cgcgaatcaccaaccacactccaattagtctagctgcttgataagatcgcggtccatgta	1200
Qy	1206	tgcttaaatgccatgcaaaagcaacagctgcttaacatgcactttaatggtctacccatc	1265
Db	1201	tgcttaaatgcatgcaaaagcaacagctgcttaacatgcactttaatggtctacccatc	1260
Qy	1266	tcaaccacacacaaacaacatgcttttcttcatcatcacacacacacccgttatata	1325
Db	1261	tcaaccacacacacaaacaacatgcttttcttcatcatcacacacacacccgttatata	1320
Qy	1326	ttcatctcttcgcgcaccccaattctcttcaacttcaacacagtcataactgcatgctgt	1385
Db	1321	ttcatctcttcgcgcaccccaattctcttcaacttcaacacagtcataactgcatgctgt	1380
Qy	1386	gtcatcccatgcccacaatctcatatgcatgattcccaaccacttctcttatataatact	1445
Db	1381	gtcatcccatgcccacaatctcatatgcatgattcccaaccacttctcttatataatact	1440
Qy	1446	ataaatcccttaatatcatctactcttcttaatcatcatcatccacagagtaactact	1505
Db	1441	ataaatcccttaatatcatctactcttcttaatcatcatcatccacagagtaactact	1500
Qy	1506	ctactactataatacccccaaccacactcatatcaatactactac	1552
Db	1501	ctactactataatacccccaaccacactcatatcaatactactac	1547
RESULT 4			
AAZ10392			
ID	AAZ10392 standard; DNA; 1244 BP.		
AC	AAZ10392;		
XX			
DT	15-NOV-1999 (first entry)		
DE	Nucleotide sequence of the bean phaseolin terminator.		
XX			
KW	Fao gene; plant metabolism; fatty acid oxidation enzyme;		
KW	polyhydroxyalkanoate; oil composition; seed production; plant biomass;		
KW	transgenic plant; terminator; bean, ss.		
XX			
OS	Phaseolus sp.		
XX			
PN	W09945122-A1.		
XX			
PD	10-SEP-1999.		
XX			
PF	05-MAR-1999; 99WO-US04999.		
XX			
PR	06-MAR-1998; 98US-0077107.		
XX			
PA	(META-) METABOLIX INC.		
XX			
PI	Boynton L, Huismann GW, Moloney M, Patterson N, Peoples OP;		
XX	Snell K;		
DR	WPI; 1999-540850/45.		

XX Modifying fatty acid metabolism in plants, useful for increasing  
PT biomass and producing specific polymers in seeds  
XX  
XX  
PS Example 3: Page 71; 79pp; English.  
XX  
CC The present sequence represents the bean phaseolin terminator, which  
CC is used to construct plasmids for the expression of the P. putida  
CC faabA gene. The gene encodes an enzyme that may be used in the  
CC method of the invention. The specification describes a method for  
CC manipulating the metabolism of a plant, and comprises expressing a  
CC heterologous gene encoding fatty acid oxidation enzymes in the  
CC cytosol or plastids other than the peroxisomes, glyoxisomes or  
CC mitochondria of the plant. The method may be used to enhance the  
CC biological production of polyhydroxyalkanoates or novel oil compositions  
CC in a transgenic plant. Plants which may be used to produce these  
CC compounds in this way include Brassicas, maize, soybean, cottonseed,  
CC sunflower, palm, coconut, safflower, peanut, mustards, flax, tobacco and  
CC alfalfa. The method may also be used to prevent or suppress seed  
CC production and therefore increase the production of biomass (leaves,  
CC stems, stalks) by plants.  
XX  
SQ Sequence 1244 BP; 456 A; 174 C; 178 G; 436 T; 0 other;

	Query Match	31.0%;	Score 1226.8;	DB 20;	Length 1244;	
	Best Local Similarity	99.4%;	Pred. No.1.4e-164;			
	Matches 1231:	Conservative	0;	Mismatches 7;	Indels 0;	Gaps 0;
OY	2720	aatcctgaagcttaataagaatgaaactaaaatgcacgtgatgltglaagagctcacatggagagc	2779			
Db	7	aatgcgaagcttaataatgaatgaaactaaatgcacgtgatgltglaagagctcacatggagagc	66			
OY	2780	atggcaatattgtatccgaccatgtaacagtataaataaacctgagctccatccatcttctc	2839			
Db	67	atggcaatattgtatccgaccatgtaacagtataaataaacctgagctccatccatcttctc	126			
OY	2840	atggaataaccaaaggatgtagtatgatatattaacacctctcatcgacacttatgtttcta	2899			
Db	127	atggaataaccaaaggatgtagtatgatatattaacacctctcatcgacacttatgtttcta	186			
OY	2900	tgtataatttcctcttatattataataatcatcvtgaacogtgtaagcgcttatggaaatgcttc	2959			
Db	187	tgtataatttcctcttatattataataatcatcvtgaacogtgtaagcgcttatggaaatgcttc	246			
OY	2960	aaatagtacaacaaaacaaatggttactatgaagctttttaaacaattcttaacttgaagcatt	3019			
Db	247	aaatagtacaacaaaacaaatggttactatgaagctttttaaacaattcttaacttgaagcatt	306			
OY	3020	gttaacagagacataaagtgtttaagaagacataacaattataatgaagaaagtttgttcoca	3079			
Db	307	gttaacagagacataaagtgtttaagaagacataacaattataatgaagaaagtttgttcoca	366			
OY	3080	tttatatattatataatlaccoccatatgtaattatataattagatgtaaggaagacataacaa	3139			
Db	367	ttttatatattatataatlaccoccatatgtaattatataattagatgtaaggaagacataacaa	426			
OY	3140	ttaataaaggagaagttttatcccatttatattatattatattatattacccaacatttatattata	3199			
Db	427	ttaataaaggagaagttttatcccatttatattatattatattatattatattacccaacatttatattata	486			
OY	3200	cctatccactattttaatgctttataaagtttgatgccatgatalatttctaatttttagt	3259			
Db	487	cctatccactattttaatgctttataaagtttgatgccatgatalatttctaatttttagt	546			
OY	3260	tgaatagtataagaagggtactatttgaacctccttactcgtataaagggttgatcat	3319			
Db	547	tgaatagtataagaagggtactatttgaacctccttactcgtataaagggttgatcat	606			
OY	3320	ccttaagaatgggctctttaattttatgttcttcacagataaaaaaaataatgaagtt	3379			
Db	607	ccttaagaatgggctctttaattttatgttcttcacagataaaaaaaataatgaagtt	666			

[illegible][illegible]





[illegible]

CC	RESULT	8
CC	AAAT03006	
CC	AAAT03006	standard; DNA, 1240 BP.
CC	AAAT03006;	
CC	13-JUN-1996	(first entry)
CC	Chymosin	open reading frame.
CC	al-3;	albino mutant; light-regulated; Neurospora; bread mould;
CC	heterologous gene;	expression; control; chymosin; ss.
CC	Mammalian	sp.
CC	W09530739-A1.	
CC	16-NOV-1995.	
CC	09-MAY-1995;	95WO-US05716.
CC	10-MAY-1994;	94US-0240372.
CC	(UYHA-) UNIV	HAWAII.
CC	Kato EK,	Stuart WD;
CC	WPI; 1995-404108/51.	
CC	Nucleic acid	for expression of heterologous protein - contains
CC	albino promoter	for light induced expression in filamentous fungi
CC	Example 2;	Fig 7; 26pp; English.
CC	The DNA	is that of a mammalian gene (open reading frame) encoding
CC	chymosin.	The gene was placed in operable linkage with the al-3
CC	promoter (see	AAAT03005). The al-3 gene controls the production of
CC	geranyl geranyl	pyrophosphatase (GGPP) synthetase. GGPP is a precursor
CC	for carotenoids	and xanthophylls. It has been shown that exposure to
CC	light	increases the transcription level of GGPP synthetase 15-45 fold.
CC	Light	activates a number of genes in the common bread mould, Neurospora.
CC	This	can be used to regulate the expression of genes encoding

CC heterologous proteins, e.g. chymosin, in recombinant production systems.  
CC use of a light-regulated promoter is a simple and effective way to  
CC control expression and allows timing to be adapted to the physiological  
CC status of the host.  
XX  
SQ Sequence 1240 BP; 274 A; 374 C; 339 G; 253 T; 0 other;

Query Match	22.78;	Score 897.4;	DB 16;	Length 1240;
Best Local Similarity	87.48;	Pred. No. 3.6e-118;		
Matches 982; Conservative	0;	Mismatches 141;	Indels 0;	Gaps 0;

OY	1609	tcgtctgctgttactcaacgcgtctcgaaatacccgcatctctctctcaaaaagttaagttc	1668
Db	100	ttcttcgcctctctcccaaggcgctcgaaataccaaagatctcctctgtataaagagcaagtctc	159
OY	1669	tcgcctaagcgctcgaaagaaacatcgagctctctgaagaactctctcgaaaacaacagtatg	1728
Db	160	tgaagaaagcgctcgaaagaaagcagtgggctctctcgaggaactctctctcgaaaacaagcagtatg	219
OY	1729	gcatacgaagaaatctaccggctctcggtgtgaagtgttgtagcgctgacatactcaactacc	1788
Db	220	gcatacgaagaaatctaccggctctcggtggaggtgtgccaagctgtgcccttgacaactacc	279
OY	1789	ttgatatgaataactcttgggaagaatactactcccgaaaccccgctccaagaagtcaacgttc	1848
Db	280	tgtatagatcgacttattgtggaaagatactactccggaaccccgcccaagagattcaacgttgc	339
OY	1849	tctttgtactcgttctctcgtactcttcgttccctctactactacgcgaagagcaatgcct	1908
Db	340	tgtttgaaactcgtccctctcgactcttcgggtlaacccctctactactacgcgaagagcaatgcct	399
OY	1909	gcaagaagccacccaagaagattctgacatccgaagaagtcgtccatctccagaagatlaagacaac	1968
Db	400	gcaaaaaacacacacgcgtcttcgacccgaagaagtgtctcaacctctccagaactctggacaagc	459
OY	1969	ccttgtctatacactacacgtgtacaggtagacatgacaagaaactctlaagctatagatacgtca	2028
Db	460	ccctgtctatacactacacgtgtacaggaacagacagatgaaaggtcactcttagtctatgacacgtca	519
OY	2029	ctgtctccacaattgtgtgaaatccacaacagaaagtagacttagaccccaagaacaagttg	2088
Db	520	ctgtctccacaattgtgtgaaatccacaacagaaagtagacttagacccaagaagaccgggg	579
OY	2089	atgtcttcaacctgaacagaaatctgcatgtgacactcttgatataagcatccatcgtctcgt	2148
Db	580	acgtcttcaacctatgcgaagaattctgacaaggatctcctggggatgtgctcaacccctcgtcgtcct	639
OY	2149	cagaagatctcgatacctgtgtgttgacaacaatgaagaaccgaactagtaagctcaagaact	2208
Db	640	cagaagatctcgatacctgtgtgttgacaacaatgaagaacagactgtgtgcccacaagcc	699
OY	2209	tgtcttcgtgtttacaatgaaagaagaaatggtccagagagaagcatatgtcaacgtcttgagactatg	2268
Db	700	tgtcttcgtgtttacaatgaaagaagaaatggtccagagagaagcatatgtcaacgtcttggggccatcgt	759
OY	2269	atccatcciaactatacagaagatctctcttcaactgggtttccaatcaactgttcagacagtatcgtg	2328
Db	760	accgcgtctactataacaaggggtccctctgactgtggtgtgcctgtgaacatgtgaacgactatggtgc	819
OY	2329	aattcaactgtgaaagatgttcaacatacgaacaggtgtgtgtgttgatcatgtgaaagtgtgaatgc	2388
Db	820	agttcaactgtgaaagatgttcaacatacgaacaggtgtgtgtgtgtgtgctgtgaggtgtgctgtgc	879
OY	2389	aagctatcttggatataccggttaagttccaagctgtgtcagacacttagacagcagacatctcaaca	2448
Db	880	agacacatctcttgacaacggtcaccttccaagctgtgtcgtgcgcagccagcacatctcctaaca	939
OY	2449	ttcagcaagctatgtggagccacacaaacaaacagttacgtgtgagtttgcataagattgtgcaca	2508
Db	940	tccagcgagccatgttggagccacacaaacaaacagttacgtgtgagtttgcatacgtacgttgcagca	999
OY	2509	accttaacatcatctccacacgtttgtcttlttgaaatcaacgtcaagatgttaccacactgacc	2568





```
RESULT 10
AA0180
ID AA0180 standard; cDNA; 1460 BP.
XX
AC AA0180;
XX
DT 25-JAN-1992 (first entry)
XX
DE Sequence of recombinant CGF4 carrying the rennin coding sequence.
XX
KM Yeast expression vector; GAL1 promoter; Saccharomyces cerevisiae;
XX
KW ss.
XX
OS Bos taurus.
XX
FT Key Location/Qualifiers
FT CDS 205..1350
FT /*tag= a
XX
PN GB2137208-A.
XX
PD 03-OCT-1984.
XX
PF 28-FEB-1984; 84GB-0405129.
XX
PR 28-FEB-1983; 83US-0470911.
XX
PA (COLB ) COLLABORATIVE RES INC.
XX
PI Botstein D, Davis RW, Fink GR, Taunton-Ridgby A, Knowlton RG;
PI Mao JT, Moir DT, Golf CG;
XX
DR WPI; 1984-245517/40.
XX
P-PSDB; AAP40218.
XX
PT DNA segment confg. GAL1 promoter linked to gene - useful for
PT direction of expression of the gene in yeast cell
XX
PS Example; Table 4, Page 21-23; 35pp; English.
XX
XX The inventors claim a DNA segment confg. GAL1 promoter linked to
CC gene - useful for direction of expression of the gene in yeast cell.
CC The recombinant material carrying a GAL1 promoter of the yeast
CC galactokinase gene may be used in expressing a desired protein, esp.
CC bovine growth hormone, interferon, prorennin or preprorennin. In the
CC yeast cell. Strains of Saccharomyces cerevisiae producing the
CC polypeptides are produced. Yeast strains deposited as ATCC 20643,
CC 20661, 20662 and 20663, strain designations CGY 196, 457, 461 and
CC 528, resp. are new.
CC
XX
SQ Sequence 1460 BP; 328 A; 440 C; 399 G; 293 T; 0 other;
XX
XX
Query Match 22.7%; Score 897.2; DB 5; Length 1460;
Best Local Similarity 87.7%; Pred. No. 3.8e-118;
Matches 980; Conservative 0; Mismatches 138; Indels 0; Gaps 0;
OY 1609 tcgttcgttactacagctgtgagatcacccgcatctctctctcaagaagtaagtc 1668
DB 233 tcttcgctctccccaagcgctgagatcaccaagatccctcgtctcaagaagttccc 292
OY 1669 tccgtaagcgctgaagaacatgactctctgaagaactctctgcaagaacaagatag 1728
DB 293 tgagggaagcgctgaagagcatggtctctggaagactctctgcaagaacaagcatag 352
OY 1729 gatacgcagcaagatctccggttcgtggaagtgtgctgagcgcaacttacaactacc 1788
DB 353 gatacgcagcaagatctccggttcgtggaagtgtgctgagcgcaacttacaactacc 412
OY 1789 ttgataagtaactcttggaagaatctactcgcgaaccccgctcaagaagttcacggttc 1848
DB 413 tggatagtaagtaactcttggaagaatctactcgcgaaccccgctcaagaagttcacggttc 472
```

```
OY 1849 tcttgatactggttctctgacttctggttccctctatctacttgcaagaagtaagtcct 1908
DB 473 tgttgacaactgctcctctgactctggttaccctctatctacttgcaagaagtaagtcct 532
OY 1909 gcaagaaccaccaagaatctgcatccggaagaagtctcactcctccaaacttaagcaaac 1968
DB 533 gcaaaaccaccagcgtctgcgaaccggaagaagtctcactcctccaaacttgcaagc 592
OY 1969 ccttgtctatacactacgtagtaagtagtaagtaagtaagtaagtaagtaagtaagtaag 2028
DB 593 cccgtctatccactacggaagcagcagcatgcaagggatctcgtgtatgacacggtca 652
OY 2029 ctgtctccaagaattgtagacatccaagagatgagatgagatgagatgagatgagatgag 2088
DB 653 ctgtctccaagaattgtagacatccaagagatgagatgagatgagatgagatgagatgag 712
OY 2089 atgtctcaactatgcaagaattcgatgcatcctgtgtatgtagtaaccactcgtcgtcgt 2148
DB 713 acgtcttcaactatgcaagaattcgatgcatcctgtgtatgtagtaaccactcgtcgtcgt 772
OY 2149 cagaagtaactgataactctgtttgacaacatgataagacacacacacacacacacacac 2208
DB 773 cagaagtaactgataactctgtttgacaacatgataagacacacacacacacacacacac 832
OY 2209 tgttctcggtttacatggaacaggaatggccaggaagacatgcatcagcttggaagtatg 2268
DB 833 tgttctcggtttacatggaacaggaatggccaggaagacatgcatcagcttggaagtatg 892
OY 2269 atccatctactacacagaagatctctcactgtgtctcagtaactgttgcaagaagtaagtc 2328
DB 893 accgtctactacacagaagatctctcactgtgtctcagtaactgttgcaagaagtaagtc 952
OY 2329 aattcaactgtggaacatgttcaacacacacacacacacacacacacacacacacacacac 2388
DB 953 agtcaactgtggaacatgttcaacacacacacacacacacacacacacacacacacacac 1012
OY 2389 aagctatcttgatataccggtacagctgtcgaagctgtcgaagctgtcgaagctgtcgaag 2448
DB 1013 aagctatcttgatataccggtacagctgtcgaagctgtcgaagctgtcgaagctgtcgaag 1072
OY 2449 ttcaagaacatgtaggaacacacacacacacacacacacacacacacacacacacacacac 2508
DB 1073 ttcaagaacatgtaggaacacacacacacacacacacacacacacacacacacacacacac 1132
OY 2509 acctgaagtaactgcttaagaagtgtcttgaagataaagcgaagtgtaaccacacacacac 2568
DB 1133 acctgaagtaactgcttaagaagtgtcttgaagataaagcgaagtgtaaccacacacacac 1192
OY 2569 ccttcgctataccacagcagatcaaggttctgcaaccagatgagatccagagtggaagac 2628
DB 1193 ccttcgctataccacagcagatcaaggttctgcaaccagatgagatccagagtggaagac 1252
OY 2629 attccagaacatgtagcttggaagatgtagtcaattcgtgtagtaactaagcgtcttgaca 2688
DB 1253 attccagaacatgtagcttggaagatgtagtcaattcgtgtagtaactaagcgtcttgaca 1312
OY 2689 gggccacaacactcgttgggtgagctgaagcaaatctga 2726
DB 1313 gggccacaacactcgttgggtgagctgaagcaaatctga 1350
RESULT 11
AA014051
ID AA014051 standard; DNA; 1210 BP.
XX
AC AA014051;
XX
DT 06-JAN-1992 (first entry)
XX
KM Rennin gene.
XX
KW Prorennin; alpha-S1-casein gene; insulin-like growth factor I; IGF-I;
mammary gland; ss.
```



```

FT XX /number= 2
PN XX US5948682-A.
PD 07-SEP-1999.
XX
XX
XX 25-APR-1997; 97US-0846021.
XX
XX 25-APR-1997; 97US-0846021.
PR 25-APR-1997; 97US-0846021.
PR 22-FEB-1991; 91US-0659835.
PR 16-NOV-1993; 93US-0142418.
PR 30-DEC-1994; 94US-0366783.
XX
XX (SEMB-) SEMBIOSYS GENETICS INC.
XX
XX Moloney MM;
XX
XX WPI; 1999-517960/43.
DR P-PSDB; AAY33830.
XX
XX Expression of a heterologous polypeptide on an oil body protein is
PT useful for the production of e.g. enzymes, antibodies, hormones
XX
XX Claim 15; Fig 6; 48bp; English.
XX
XX This is the nucleotide sequence of a HindIII fragment containing the
CC oleosin-spacer-Met-prothimosin sequence. This HindIII fragment was
CC joined to a nopaline synthase terminator and cloned into binary vector
CC pCNI1559. The resulting plasmid was called pBSORPNT and introduced
CC into A. tumefaciens. The resulting bacterial strain was used to transform
CC B. napus plants.
CC The DNA which encodes a chimeric fusion protein that consists of the
CC oil-body targeting sequence, a transcription regulation sequence and the
CC DNA of the protein of interest can be used to produce antibodies,
CC glycanases, hormones, proteases, protease inhibitors, seed storage
CC proteins, thrombin inhibitors, hirudin, interleukins, chymosin, cystatin,
CC xylenase, carp growth hormone, zein or a collagenase.
CC The enzyme may be cleaved from the oil body protein or used in
CC association with the oil body fraction.
CC Allows production of commercially important proteins on a superior scale
CC to production by conventional systems. The expressed heterologous
CC protein can be easily separated from host cell components
CC due to the use of the oil body as a carrier protein.
XX
XX Sequence 2733 BP; 699 A; 681 C; 684 G; 669 T; 0 other;
SQ
Query Match 22.6%; Score 895.8; DB 20; Length 2733;
Best Local Similarity 86.4%; Pred. No. 5.8e-118;
Matches 972; Conservative 0; Mismatches 127; Indels 0; Gaps 0;
QY 1629 gctagatcaccgcgcctctctctacaaagtgatctcccgtaagcgctgaagac 1688
    |||||||
DB 1630 gctagatcaccgcgcctctctctacaaagtgatctcccgtaagcgctgaagag 1689
QY 1689 catggaactctagaagactctctgcagaacaacagtaatggcatcagcaagaatctacc 1748
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1690 catggctctcggaggaactctctgcagaacaacagtaatggcatcagcaagaatctacc 1749
QY 1749 ggcttcggtgaagtgtgtagcggtccacttaccaactactgttaagtaactttggg 1808
    ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1750 ggcttcggtgaagtgtgtagcggtccacttaccaactactgttaagtaactttggg 1809
QY 1809 aagatctaacctcgggaacccgcctcaagaagttcacggtctctttagatcagttgctcct 1868
    ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1810 aagatctaacctcgggaacccgcctcaagaagttcacggtctctttagatcagttgctcct 1869
QY 1869 gactctgggtccctctatatctactgcaagcaaatgctctcaagaacacacaaagattc 1928
    ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1870 gactctgggtccctctatatctactgcaagcaaatgctctcaagaacacacacagctc 1929
QY 1929 gatccgagaagtcgtccacttcagaacttaggcaaaccttgcctatatacactaaggt 1988
    || ||||||| ||||||| ||||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

DB 1930 gaccggaagaagtcgtccacttcacagaacctgggcgaagccctgtctatccactacggg 1989
QY 1989 acaagtagcagtcgaaggaatcttagctatagatccgltcaactgtctccaactgttgac 2048
    ||||| ||||||| ||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
DB 1990 acaagcagacatgcagagatcccttggtctatgacacgcgtcaactgtctccaactgttgac 2049
QY 2049 attcaacagacatgagacttagaccaccaagaacaaagtgatgtcttccaactatgcagaa 2108
    || ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2050 atccagcagacatgagcttagcccttagaccaccaagagcccggtgagcttccaactatgcagaa 2109
QY 2109 ttcatgtgcatcctctgtatgcatlaccatcgtctgcgtgtaagagtactcgataccggtg 2168
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2110 ttgcacgagatccctgggatgtacccttcgtctgcgtcctcctagatgactgataccggtg 2169
QY 2169 tttagacaacatgataagccacccctagctagctcaagaactgttctcgtttatcatgac 2228
    ||||||| ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2170 tttagacaacatgataagccacccctagctagctcaagaactgttctcgtttatcatgac 2229
QY 2229 aggaatggccagagagacatgtctacgcttggagactatgttccatccactactacacagaa 2288
    ||||||| ||||||| ||||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2230 aggaatggccagagagacatgtctacgcttggagactatgttccatccactactacacagaa 2289
QY 2289 tctctcaactgggttccaagtctgtgcagcagactgtgcaattcaactgttggagacgttc 2348
    || ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2290 tccctgcacttgggtgcccgtgtacagtgacagactgtgcagactgttgcactgttggagacgttc 2349
QY 2349 accatcagcgtgtggtgtgtgtgcatggaagtgatgtacgaactgttgcactgttgcactgt 2408
    ||||||| ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2350 accatcagcgtgtggtgtgtgtgcatggaagtgatgtacgaactgttgcactgttgcactgt 2409
QY 2409 acgtccaagcgtgtgcagacacttagcagcagacatctccaacatlcagcaagctattgagcc 2468
    || ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2410 accatcagcgtgtgcagacacttagcagcagacatctccaacatlcagcaagcattgagcc 2469
QY 2469 acaagaacccgactagctgtgattgacatagattggcagaacacttagctaatgtctaca 2528
    || ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2470 acaagaacccgactagctgtgattgacatagattggcagaacacttagctaatgtctaca 2529
QY 2529 gttgtcttggatcaacagcgaagatgtaccacactgtacccttcgcgtatataccagcag 2588
    || ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2530 gttgtcttggatcaacagcgaagatgtaccacactgtacccttcgcgtatataccagcag 2589
QY 2589 gatcaagggtctcgcacacagtgatctccagagtgaagaacatctccagaagaatgatacttg 2648
    || || ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2590 gaccagggctcttgacacagcgtgtccagagtgaaatacatctccagaagaatgatacttg 2649
QY 2649 ggaagtgtgtcattctgtgtaactactacagcgtctttagcaagggccaacaactgtttggg 2708
    || ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2650 gggagtgtttcatccgcagagatlaacagcgtctttagcaagggccaacaactgtttggg 2709
QY 2709 ctgactaaagcaatctgaa 2727
    || || |||||||
DB 2710 ctggcacaagcacttgaa 2728

```

RESULT 13  
 AAN40055  
 ID AAN40055 standard; DNA; 1278 BP.  
 XX  
 XX AAN40055:  
 AC  
 AC 02-FEB-1992 (first entry)  
 DE  
 DE Sequence of prothimosin gene.  
 XX  
 XX Prothimosin expression vector; E. coli trp operon; chymosin; ss.  
 OS  
 OS Bos taurus.  
 XX  
 XX Key location/Qualifiers  
 FT CDS 3..1130  
 FT /\*tag= a  
 FT poly\_a-signal 1245..1250

FT XX /\*tag= b  
 PN XX EPI21775-A.  
 XX 17-OCT-1984.  
 PD 07-MAR-1984; 84EP-0102451.  
 XX 09-MAR-1983; 83JP-0038439.  
 PR (BEP/) BEPU T.  
 PA Beppu T, Uozumi T, Nishimori K, Shimizu N, Kawaguchi Y;  
 PI Hidaka M;  
 DR WPI: 1984-258001/42.  
 DR P-PSDB: AAP40078.  
 XX Expression plasmid comprising prochymosin gene and vector -  
 PT useful for transforming Escherichia coli for prochymosin prodn.  
 XX Disclosure: Fig 1; 59pp; English.  
 XX The inventors claim the prochymosin gene comprising a nucleotide  
 CC sequence from (a) the 1st codon (GCT) to the 365th codon (ATC), or  
 CC (b) the 5th codon (CGG) to the 365th codon (ATC); and recombinant  
 CC plasmids harboured by Escherichia coli strains deposited as FERM BP-  
 CC 262, -263 and -264. Any portion of the nucleotide sequence as  
 CC described in AAN40055 can be used. Also claimed is a vector derived  
 CC from plasmid pBR322. Typically plasmid pBR301 is obtd. from pOCT 2.  
 CC The transcriptional direction of pOCT 3 is opposite to that of  
 CC pOCT 2; it is clockwise in pOCT 2 (5' to 3') whereas counter  
 CC clockwise in pOCT 3.  
 XX  
 SQ Sequence 1278 BP: 309 A; 384 C; 326 G; 259 T; 0 other:  
 Query Match 22.6%; Score 895.6; DB 5; Length 1278;  
 Best Local Similarity 87.6%; Pred. No. 6.4e-118;  
 Matches 979; Conservative 0; Mismatches 139; Indels 0; Gaps 0;  
 QY 1609 tcgtgtgttcttaccagctctgagatcaccgcatctctcttcaagaagtaagctc 1668  
 DB 13 tcttcgctctctccagggcgctgagatccagagatccctctgtacaaggaagctc 72  
 QY 1669 tccgtaaggcgctgaaggaatgactcttgaagaactcttctcagaagaacagatg 1728  
 DB 73 tgaagaaaggcgctgaaggaagcattggtctcggaggaactcttcgaagaacagcagatg 132  
 QY 1729 gcatcagcagaagctactccggtctcggtgaagttgtagctgacacttcaactacc 1788  
 DB 133 gcatcagcagaagctactccggtctcggtgaggtgagcaggtgctccctgacactacc 192  
 QY 1789 ttgtagtcaatctcttgggaagatctaccctcgaaaccgcccctcaagaagttcacgctc 1848  
 DB 193 tggatagctactcttgggaagatctaccctcgaaaccgcccctcaagaagttcacgctc 252  
 QY 1849 tcttgatagctggtctctactcttgggtctccctctatctactgcaaggaagctgct 1908  
 DB 253 tgttgacactggtctctcttgggtctccctctatctactgcaaggaagctgct 312  
 QY 1909 gcaagaacacacaaagatctgacccgaagaagtcgtccacacttcaagaactaggaac 1968  
 DB 313 gcaaaacacacagcgtctgacccgaagaagtcgtccacacttcaagaactggaagc 372  
 QY 1969 cctgtctatatacactagtgatcagtgatcagaagaagtaactttagctatgatacgtca 2028  
 DB 373 cccgtctatatacactagtgatcagaagaagtaactttagctatgatacgtca 432  
 QY 2029 ctgctccaacatgtgatactcaacagagctaggaacttagcaccacaaacaggtg 2088  
 DB 433 ctgctccaacatgtgatactcaacagagctaggaacttagcaccacaaacaggtg 492

QY 2089 atgtctcactatgcagaatctcgaatggcactctgtgtagtgcataccatccgtcgt 2148  
 DB 493 agctctcactatgcagaatctcgaatggcactctgtgtagtgcataccatccgtcgtc 552  
 QY 2149 cagagctactgatactctgttggacaacatgataacccagacacttagtgcagact 2208  
 DB 553 cagagctactgatactctgttggacaacatgataacccagacacttagtgcagact 612  
 QY 2209 tgtctcggtttacatggagcaggaatggcaggaagaagatgtcaagctttagatg 2268  
 DB 613 tgtctcggtttacatggagcaggaatggcaggaagaagatgtcaagctttagatg 672  
 QY 2269 atccatctactacacagaatctctcactgtgttccagctactgtgacagagtagtgc 2328  
 DB 673 accgctctactacacagaatctctcactgtgttccagctactgtgacagagtagtgc 732  
 QY 2329 aattcactgtgacagatgtcaacatcagcgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2388  
 DB 733 agttcactgtgacagatgtcaacatcagcgtgtgtgtgtgtgtgtgtgtgtgtgtgt 792  
 QY 2389 aagctatcttgatccggtgtcagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2448  
 DB 793 aggcatacttgatccggtgtcagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 852  
 QY 2449 ttcagaagctatgtgagcacaacagaacagtagcgtgtgtgtgtgtgtgtgtgtgtgt 2508  
 DB 853 ttcagaagctatgtgagcacaacagaacagtagcgtgtgtgtgtgtgtgtgtgtgtgt 912  
 QY 2509 acctagctatagctcctacagttgtcttggatcaacgaagatgtaccacatgacc 2568  
 DB 913 acctagctatagctcctacagttgtcttggatcaacgaagatgtaccacatgacc 972  
 QY 2569 cctcgcctatatacagcaggaatcaaggtgtctgacaggtgtgtgtgtgtgtgtgtgtgt 2628  
 DB 973 cctcgcctatatacagcaggaatcaaggtgtctgacaggtgtgtgtgtgtgtgtgtgtgt 1032  
 QY 2629 attccagaatgtgattcttggagatgtgttcattctgtgtagtactacagcgtcttggaca 2688  
 DB 1033 attccagaatgtgattcttggagatgtgttcattctgtgtagtactacagcgtcttggaca 1092  
 QY 2689 gggccaacaacctgt 2726  
 DB 1093 gggccaacaacctgt 1130  
 RESULT 14  
 ID AAN20043 standard; DNA; 1460 BP.  
 XX AAN20043;  
 AC AAN20043;  
 XX 16-DEC-1992 (first entry)  
 DE Pre-prorennin-A gene DNA sequence.  
 XX Pre-pro-rennin; rennin; prorennin; enzyme; EC-3.4.23.4; chymosin;  
 KW protease; milk-clotting enzyme; ss.  
 XX Bos taurus.  
 OS  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 205..1350 /\*tag= a  
 FT /label= /note= "pre-prorennin-A gene"  
 FT  
 XX GB2091271-A.  
 XX 28-JUL-1982.  
 PD 15-JAN-1982; 82GB-0001120.  
 XX 01-DEC-1981; 81US-0325481.

PR 16-JAN-1981; 81US-0225717.  
XX  
PA (COLB ) COLLABORATIVE RES INC.  
XX  
PI Alford BL, Mao J, Moir DT;  
XX  
XX WPI; 1982-62028E/30 (62028E) .  
DR P-PSDB; AAP20038.  
XX  
PT Transformed cells producing rennin and its precursors - contg.  
PT appropriate recombinant DNA material  
XX  
XX Disclosure: Table 1: 39pp; English.  
PS  
XX  
XX Bases 1-204 and 1351 to 1460 are attached to the pre-prorennin but  
CC can be removed and are not essential to use of the gene in  
CC expression. The gene may be ligated into plasmid pCGB21 and  
CC expressed in E. coli. The resulting expressed enzyme is a well  
CC known milk-clotting enzyme used in cheese-making.  
XX  
SQ Sequence 1460 BP; 329 A; 440 C; 397 G; 294 T; 0 other;

CC The Kluveromyces expression systems provide highly efficient secretion  
CC and processing of a wide variety of proteins.  
CC Sequences identical to those published in EP301669.  
XX  
SQ Sequence 2727 BP; 752 A; 654 C; 572 G; 749 T; 0 other;

```
Query Match      22.6%: Score 895.6; DB 10; Length 2727;
Best Local Similarity 82.7%: Pred. No. 6.2e-118;
Matches 1024; Conservative 0; Mismatches 214; Indels 0; Gaps 0;

QY 1610 cgtgtcgttactcaagctgtgagatcaaccgcgacatctctctacaagaagtaagtctct 1669
    ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 669 cgatgcttcccatcatagatgagatcaccaagatccctctgtacaagaagtaagtctct 728

QY 1670 ccgtaagcgcttgaaagaaacttgactctctgaagaactctctgcagaacaacatagtg 1729
    | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 729 gaggaagcgcttgaaagaaacttgactctctgaagaactctctgcagaacaacatagtg 788

QY 1730 catcagcagaagtaactcgcgcttcgtgaaagtgtctagcgtgcaccattaccaactact 1789
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 789 catcagcagaagtaactcgcgcttcgtgaaagtgtctagcgtgcaccattaccaactact 848

QY 1790 tgatagtaactacttgggaagatctactcggaaacccgcctcaagaagttcacggttct 1849
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 849 ggaagcagtaacttgggaagatctactcggaaacccgcctcaagaagttcacggttct 908

QY 1850 cttgtatctgttccctctgtctctgtctgtccctctactactatgcagaagatgctg 1909
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 909 gtttgaacatgtgctctctgtctctgtctgtgtccctctactactatgcagaagatgctg 968

QY 1910 caagaacccaagaagatctcgtccagaagaagtcgtccacctccagaacttaggcaaac 1969
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 969 caaaacccaagaagatctcgtccagaagaagtcgtccacctccagaacttaggcaaac 1028

QY 1970 ctgtctatacactaaggtacaggtagcatgcagaagaactcttaggtctatgataccgtcac 2029
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1029 cctgtctatacactaaggtacaggtagcatgcagaagaactcttaggtctatgataccgtcac 1088

QY 2030 tgtctccaacatgttgagacatccaagacgttagacttagcaccagaacagagcggtga 2089
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1089 tgtctccaacatgttgagacatccaagacgttagacttagcaccagaacagagcggtga 1148

QY 2090 tgtctccactatgcagaatctcgtatgcatccttgttatgcataccatccatcgctcgctc 2149
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1149 cgtctccactatgcagaatctcgtatgcatccttgttatgcataccatccatccatcgctcgctc 1208

QY 2150 agagtaactgataactgtgttgtaacaacatgataaccgaacactagtaactcaagaact 2209
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1209 agagtaactgataactgtgttgtaacaacatgataaccgaacactagtaactcaagaact 1268

QY 2210 gttctcggttatacatgagcaagaagatgagcaagagacatgctcaagcttgagactatgta 2269
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1269 gttctcggttatacatgagcaagaagatgagcaagagacatgctcaagcttgagactatgta 1328

QY 2270 tccatccactactacaagaatctcttcaactgggtccacgtacactgtgcagcagtaactgga 2329
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1329 cccgtccactactacaagaatctcttcaactgggtccacgtacactgtgcagcagtaactgga 1388

QY 2330 attcaactgtgacaggtgtcacacatcagcgtgtggtgtgtgcatgtgaagtgatgtca 2389
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1389 gttcaactgtgacaggtgtcacacatcagcgtgtggtgtgtgcatgtgaagtgatgtca 1448

QY 2390 agctatcttgataccggttaagtcacaaagctgtgcagactagcagcagacattctcaaat 2449
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1449 ggcacatcttgataccggttaagtcacaaagctgtgcagactagcagcagacattctcaaat 1508

QY 2450 tcagcaagctatgtgagcacaacagaacacagtacgtgagtttgacatagattgacagaa 2509
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1509 ccagcagcgcatgttgagcacaacagaacacagtacgtgagtttgacatagattgacagaa 1568

QY 2510 ccttagtaactgactcgtctgttgagatacaagcgcaagaatgtaccactgaaccc 2569
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
DB 1569 cctgagctacatgcccactgtgtcttgaagatcaatgcaaaatgtaaccactgaccc 1628
QY 2570 ctccgctataccagccagagatcaagggttctgtccacagtgattccagagtgagaacca 2629
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1629 ctccgctataccagccagagatcaagggttctgtccacagtgattccagagtgagaacca 1688

QY 2630 ttccagaagaatggaatcttgagagatgtgtcatctcgttgagtactacaacgctcttgacag 2689
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1689 ttccagaagaatggaatcttgagagatgtgtcatctcgttgagtactacaacgctcttgacag 1748

QY 2690 ggcacaacacctgtgtggtctagctaaagcaactcgtgaagcttaataagtatgaactaaa 2749
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1749 ggcacaacacctgtgtggtctagctaaagcaactcgtgaagcttaataagtatgaactaaa 1808

QY 2750 tgcattgagtggtgaagagctcagtgagagacatggaatatgtatccgacatgtaacagt 2809
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1809 ccaaggttaagtgaaattacttacttaagaagctgtgacatcttaanaaatttcttttatagc 1868

QY 2810 ataatactgagctccatctcactcttctatgataa 2847
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1869 ttatgacttagttccaattatataactatttaata 1906
```

Search completed: July 31, 2002, 21:12:36  
Job time: 25787 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 15:44:34 ; Search time 5848.1 Seconds  
(without alignments)  
9132.445 Million cell updates/sec

Title: US-09-643-755b-3  
Perfect score: 3957  
Sequence: 1 ccgcaggaatcatgtact.....accctaccactaaggtacc 3957

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST.\*  
1: em\_estdb:\*  
2: em\_esthm:\*  
3: em\_estln:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hlc:\*  
9: gb\_estl:\*  
10: gb\_estc:\*  
11: gb\_hlc:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vtc:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	387.2	9.8	472	10	BG938320 1AB015E12
2	358	9.0	631	9	A1892188 m18c05.y
3	354.8	9.0	430	10	BG937697 1AB005E01
4	337.2	8.5	727	9	A1326975 m18e03.x
5	333.8	8.2	399	10	BG937723 1AB005D06
6	301.4	7.6	1347	11	AK004109 Mus muscu
7	300.6	7.6	383	10	BG938086 1AB011A08
8	262	6.6	479	9	AA028632 m14c08.r
9	249.6	6.3	505	9	A1479358 tm27e07.x
10	246	6.2	546	9	A1324867 m18b01.x
11	242.4	6.1	517	9	A1385490 m18e03.y
12	236.2	6.0	1385	11	AK008959 Mus muscu
13	236	6.0	619	9	AM868692 MRL-SN06
14	227.4	5.7	1388	11	AK008886 Mus muscu
15	226.8	5.7	414	9	A1322423 m18b01.y
16	217.8	5.5	636	9	AM868716 MRL-SN06
17	206.4	5.2	814	10	B1761345 603043918

18	205.8	5.2	583	10	BE841742	BE841742 MRL-SN06
19	204.4	5.2	910	10	BF299798	BF299798 602030421
20	202.8	5.1	622	10	BF373831	BF373831 MRL-SN00
21	201	5.1	716	9	AM012992	AM012992 SSERN01SK
22	199.4	5.0	603	10	BE841740	BE841740 MRL-SN06
23	199	5.0	702	9	AM867433	AM867433 MRL-SN03
24	198.6	5.0	551	10	BE841714	BE841714 MRL-SN06
25	196.4	5.0	547	10	BM069200	BM069200 id171a02.y
26	196.4	5.0	548	10	BM053683	BM053683 id69b07.y
27	190.6	4.8	391	9	AA969042	AA969042 op43d07.s
28	190	4.8	433	10	W34201	W34201 ma99h09.r1
29	190	4.8	713	10	BF121716	BF121716 601759230
30	187.4	4.7	422	10	W10274	W10274 ma37g09.r1
31	185	4.7	666	10	BE841761	BE841761 MRL-SN06
32	182.6	4.6	594	9	AM863768	AM863768 MRL-SN01
33	182.2	4.6	578	10	BE840800	BE840800 MRL-SN00
34	182	4.6	564	9	AM868711	AM868711 MRL-SN06
35	181	4.6	736	10	B1821121	B1821121 603035596
36	180.8	4.6	847	10	B1490133	B1490133 603031966
37	180.8	4.6	616	10	BE841758	BE841758 MRL-SN06
38	180.4	4.6	547	10	BE841657	BE841657 MRL-SN06
39	180.4	4.6	576	9	AM868673	AM868673 MRL-SN06
40	179.6	4.5	960	10	B1517784	B1517784 603042031
41	179.4	4.5	751	9	A1076336	A1076336 oz04a06.x
42	178.8	4.5	815	10	B1818940	B1818940 603037434
43	176.4	4.5	666	9	AM959843	AM959843 EST372014
44	176.2	4.5	847	10	B1762154	B1762154 603049340
45	176	4.4	572	10	BF373812	BF373812 MRL-SN00

## ALIGNMENTS

RESULT 1  
LOCUS BG938320 472 bp mRNA linear EST 11-JUN-2001  
DEFINITION 1AB015E12 Bovine Abomasum cDNA Library Bos taurus cDNA 5', mRNA  
sequence.  
ACCESSION BG938320.1 GI:14337692  
VERSION BG938320.1  
KEYWORDS EST.  
SOURCE Bos taurus  
ORGANISM Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Kuminantia; Pecora; Bovidae;  
Bovidae; Bovinae; Bos.  
REFERENCE 1 (bases 1 to 472)  
Moore,S.S., Hansen,C., Li,C., Fu,A., Meng,Y. and Li,G.  
CDNA's from bovine abomasum tissue  
unpublished (2001)  
JOURNAL  
TITLE  
AUTHORS  
COMMENT  
Contact: Dr. Stephen Moore  
. Beef Genomics Laboratory  
Dept of AFNS, University of Alberta  
410 Agri/For, Dept of AFNS, U of A, Edmonton, AB, T6G 2P5, Canada  
Tel: 780 492 0169  
Fax: 780 492 4265  
Email: smoores@afns.ualberta.ca  
The sequence best matches gb:BOVCHYMOA (bovine chymosin a (rennin)  
mRNA) in main database at high score of 928.0 and E-value of 0.0  
PCR Primers  
FORWARD: M13 Forward  
BACKWARD: M13 Reverse  
Seq primer: T3 primer  
High quality sequence stop: 472  
POLYA-No.  
FEATURES  
source location/Qualifiers  
1..472  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/clone\_lib="Bovine Abomasum cDNA Library"  
/sex="Two males and one female mixed"  
/tissue\_type="Gastrointestinal tissue (GIT)"  
/cell\_type="Epithelial"



DB	541	AAATTGGGCGTTACAGTGCAGACAGATCATCATCATGTGAAGTGGTTCGTCACAG	600
QY	2381	tggtatgtcaagctatcttggataccggtac	2410
Db	601	TGGTCGCCCTGGCGCTTCGACACAGGAC	630
RESULT	3		
LOCUS	BG937697	430 bp	mRNA
DEFINITION	1A0005E01	Bovine Abomasum cDNA Library	Bos taurus cDNA 5', mRNA
ACCESSION	BG937697		
VERSION	BG937697.1	GI:14337069	
KEYWORDS	EST.		
SOURCE	COW.		
ORGANISM	Bos taurus		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;		
	Bovidae; Bovinae; Bos.		
	1 (bases 1 to 430)		
REFERENCE	Moore,S.S., Hansen,C., Li,C., Fu,A., Meng,Y. and Li,G.		
AUTHORS	CDNA's from bovine abomasum tissue		
TITLE	Unpublished (2001)		
JOURNAL	Contact: Dr. Stephen Moore		
COMMENT	. Beef Genomics Laboratory		
	Dept of AFNS, University of Alberta		
	410 Agri/For. Dept of AFNS, U of A, Edmonton, AB, T6G 2P5, Canada		
	Tel: 780 492 0169		
	Fax: 780 492 4265		
	Email: smoores@afns.ualberta.ca		
	The sequence best matches gb:BOVCHYMOA (bovine chymosin a (rennin)		
	mRNA) in main database at high score of 844.0 and E-value of 0.0		
	PCR primers		
	FORWARD: M13 Forward		
	BACKWARD: M13 Reverse		
	Seq primer: T3 primer		
	High quality sequence stop: 430		
	POLYA-NO.		
FEATURES			
source	Location/Qualifiers		
	1..430		
	/organism="Bos taurus"		
	/db_xref="taxon:9913"		
	/clone_lib="Bovine Abomasum cDNA Library"		
	/sex="Two males and one female mixed"		
	/tissue_type="Gastrointestinal tissue (GIT)"		
	/cell_type="Epithelial"		
	/dev_stage="Young adult"		
	/lab_host="XLI-BlueMRF"-strain"		
	/note="Organ: Abomasum; Vector: Uni-22APXR; Site_1: EcoR		
	I; Site_2: Xho I"		
BASE COUNT	97 a 130 c 119 g 84 t		
ORIGIN			
Query Match	9.0%; Score 354.8; DB 10; Length 430;		
Best Local Similarity	89.1%; Pred. No. 1.9e-36;		
Matches 383; Conservative	0; Mismatches 47; Indels 0; Gaps 0;		
QY	2158	cgataccgtgtttggaacacatgatgaacccgacacctagtagtaagaactgtgtctgg	2217
Db	1	CGATACCCGCTGTTGACACATGATGAACAGGACCTGGTGCCCAAGACCTGTTCTCGG	60
QY	2218	tttataatgacaggaatggccagagagacatgctcaagcttgagactttaagctatcatct	2277
Db	61	TTTACATGACACGAGATGGCCAGAGAGCATGCTACGCTGGGGCCATCGACCCGTCCT	120
QY	2278	actacacagatctcttacttgggttccagtcagtcgtgcagagacttgcgaatcactc	2337
Db	121	ACTACACAGAGGTCCTGCACCTGGGTCGCCGTCACAGTGCAGCAGTACTGACGCTTCAC	180
QY	2338	tggacagtgtcaacatcagcggt	2397

Db	181	TGGACAGTGTCAACATCAGCGGCTGTGTGTGTGTGTGCCTGTGAGGGTGCGCTGTCAAGGCCATCC	240
Oy	2398	tggataccggtatcgtccaaagctgtgcctgacctaagcagcgacattctcaacattcagcaag	2457
Db	241	TGGACAGCGGACCTTCACAAAGCTGTCTCGGGCCAGCAGCGACATVCTTCAAATCCAGCAG	300
Oy	2458	ctattggagccacaagaaccagtacggttgatttggacatagatttcgcgaacacctagct	2517
Db	301	CCATTGGAGCCACACAGAACCAGTACGATGACTTTGACATCATCACGCGCAACAACCTGAGCT	360
Oy	2518	aactgcctcaaatgtctctttgatgacaaaggaaatbtaccactatgacctccctgcct	2577
Db	361	AATATCCCACTGTGGTCTTTTGAGATCAATGCGAATAATGTACCACCTGACCCCTCGGCT	420
Oy	2578	ataccagcca 2587	
Db	421	ATACCAGCCA 430	
RESULT	4		
LOCUS	A1326975/c	727 bp	mRNA linear EST 23-DEC-1998
DEFINITION	mbl1e03.xl Soares mouse p3NMf19.5 Mus musculus cDNA clone IMAGE:482524 3' similar to TR.Q28075 Q28075 CHYMOSIN C.; ; mRNA sequence.		
ACCESSION	A1326975		
VERSION	A1326975		
KEYWORDS	A1326975.1 GI:4061404		
SOURCE	EST.		
ORGANISM	house mouse. Mus musculus.		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 727) Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R. The WashU-HMI Mouse EST Project Unpublished (1996) Contact: Marra M/Mouse EST Project WashU-HMI Mouse EST Project Washington University School of Medicine 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mousestewatson.wustl.edu This clone is available royalty-free through LNLN; contact the IMAGE Consortium (info@image.llnl.gov) for further information. MGI:293268 This clone was previously sequenced on the 5' end only, this new data is from the 3' end Possible reversed clone: similarity on wrong strand High quality sequence stop: 446. Location/Qualifiers 1..727 /organism="Mus musculus" /db_xref="taxon:10090" /clone="IMAGE:482524" /clone_id="Soares mouse p3NMf19.5" /dev_stage="19.5 dpc total fetus" /lab_host="DH10B (ampicillin resistant)" /note="Vector: pUT73D (Pharmacia) with a modified polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5, TGTATCCAAATCTGAAGTGGAGCGCGCATATTTTTTTTTTTTTTTT 3' ], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pUT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M.Felima Bonaldi. RNA was kindly provided by Dr. Minoru Ko (Wayne State University)."		
TITLE	JOURNAL		
COMMENT			
FEATURES	source		





Db	700	AAAGGACAGCATGCTGATGCTAGTGGAGTGGATCCCTCTACTACCATGAGGAGAGCTTAC	759
Qy	2238	tggatcagcagcactctgctgagcagcagctactgtcaaatcactctgtggacagtgctaccatcagc	2357
Db	760	TGGGTACCAAGTGTCCAAACCCAGCTACAGGCAATTAAGCTGTGGATGACATCTCCATGTAAAT	819
Qy	2358	ggtgtggtgtgtgcatgtgaagtgatgtcaagctatcttggataccggtacgctcaag	2417
Db	820	GGGAGAGCATTTGCTGCTGTGATGTGTGGCTGCGCCAAAGTATTATGACACAGGAGACCTCCTTG	879
Qy	2418	ctgtctgcagccttagcagcagcagctcttcacatcgaacgaagctatgtggacacacagaac	2477
Db	880	CTGACCGGCCCCCAAGCTCCATCGTTAAACATCCAGATCTAATGTGGTGCACAAGGCTTCT	939
Qy	2478	cagtaacggttgagtttgatgatattgtgcacaaccttagctacatgagccacagtgctctt	2537
Db	940	GGTAGCGGCGAGTACTTCTCAAGTGTACACCATCAACACCTGCTGTGATATTGTCTTC	999
Qy	2538	gagatcaacagcgaagatgtaccacactgtgacccctccgctcctataccagccagctacaag	2597
Db	1000	ACCATCGGCACATGTTTACTACCTACCAAGTGCACAGCAATGCTCATATCCGAAGATCGGTCA	1059
Qy	2598	ttctgcaccagtgatgtattccagaagtgagac-----cattccagaatg	2642
Db	1060	CACAAATTCCAGAGAGCAMACTTTGAGGAGGCGCATGTGATGCCATCAGACCTGAGATGTGG	1119
Qy	2643	atcttggagagatgtgttattatcctgtgagctacacagcgtctttgacagggccaacacctc	2702
Db	1120	GTGCTGGGGGATGTCTTCTCTGAGCTGTATTATTCACCGCTGTTTGATCGGGCAATATACAG	1179
Qy	2703	gttgggctagctaaagcaatctgaaagcttaa	2733
Db	1180	ATTGCTGTGGCTCTCTGTCATGAGTGTGA	1210
RESULT 7			
RG938086		383 bp	mRNA
LOCUS	1A0011A08	Bovine Abomasum cDNA Library	Bos taurus cDNA 5', mRNA
DEFINITION			sequence.
ACCESSION	RG938086		
VERSION	RG938086.1	GI:14337458	
KEYWORDS			EST.
SOURCE			
ORGANISM			Cow.
REFERENCE			
AUTHORS			Bos taurus
TITLE			Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
JOURNAL			Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
COMMENT			Bovidae; Bovinae; Bos.
			1 (bases 1 to 383)
			Moore,S.S., Hansen,C., Li,C., Fu,A., Meng,Y. and Li,G.
			cDNA's from bovine abomasum tissue
			unpublished (2001)
			Contact: Dr. Stephen Moore
			. Beef Genomics Laboratory
			Dept of AFNS, University of Alberta
			410 Agri/FOR, Dept of AFNS, U of A, Edmonton, AB, T6G 2P5, Canada
			Tel: 780 492 0169
			Fax: 780 492 4265
			Email: smoores@afns.ualberta.ca
			The sequence best matches gb:BPUI9786 (Bos primigenius prochimosis
			mRNA, complete cds) in main database at high score of 735.0 and
			E-value of 0.0
			PCR Primers
			FORWARD: M13 Forward
			BACKWARD: M13 Reverse
			Seq primer: T3 primer
			High quality sequence stop: 383
			POLYA-No.
FEATURES			
Source			Location/Qualifiers
			1..383
			/organism="Bos taurus"
			/db_xref="taxon:9913"

Query Match	7.6%;	Score 300.6;	DB 10;	Length 383;
Best Local Similarity	88.3%;	Pred. No. 1.8e-29;		
Matches 338;	Conservative 0;	Mismatches 44;	Indels 1;	Gaps 1;
Oy 2152	agtactcgatcacctgtgtttgacaacaatgatgaacccagacactagtagctcaagacttgt 2211			
Db 1	AGTACTCGATCACCGCTGTTTGACAACATGATGAAAGCAGCACCCTGGTGGCCCAAGACTGT 60			
Oy 2212	tctcggtttacatgtagcaagaaatggccaggaagacatgctcaagcttggagctatgtc 2271			
Db 61	TCTCGGTTTACATGAGCAGGAATGGCCAGAGAGCATCTCACGCTGGGGGCCAATCGACC 120			
Oy 2272	catcctactaacagagatctcttcaactgtgttccagtcagtcgltgagcaagtaactggcaat 2331			
Db 121	CGTCTTACTATACACAGAGGTCCTCGCACTGGGTGCCCGTATACATGACAGCATGTGGCACT 180			
Oy 2332	tcaactgtgacagtgatcaacatcaacagcggtgtgtgttctgtagctgtgaaggtgatgtcaag 2391			
Db 161	TCATCTGTGACAGTGTCAACATCACAGCGGTGTGTGTGTGTGGCTGTAGAGGTGGCTGTGAGG 240			
Oy 2392	ctatcttggataccggtacgcgtccaaagctgtgtctgcagacctagacagcagcatcttcaacatlc 2451			
Db 241	CCATCCTCGACACGGGACACCTCCAGCTGCTGGCCGCCAGAGCAGCACAATCTCAACATCC 300			
Oy 2452	agcaagcgtataggagcccaacaaacacagctggtgagtttgatcatc-agattgcgacaac 2510			
Db 301	ACCAGGCGCATTTGGAGCCACACAGAAACCGATACGGTGTGATGTGACATCTGACTGGACACAC 360			
Oy 2511	cttagctacatgacctacagttgt 2533			
Db 361	CTGAGCTACTATGCCCACTGTGCT 383			
RESULT 8				
AA028632	479 bp	mrna	linear	EST 11-SEP-1996
LOCUS	m114c08.r1	Soares mouse p3MNF19.5	mus musculus	cDNA clone
DEFINITION	IMAGE:463502.5	similar to SW:CHYM_BOVIN P00794	PROCHYMOSIN A AND B	
PRECURSORS		;	mrna	sequence.
ACCESSION	AA028632			
VERSION	AA028632.1	GI:1494769		
KEYWORDS	EST.			
SOURCE	house mouse.			
ORGANISM	Mus musculus			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
AUTHORS	1 (bases 1 to 479)			
	Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisler,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.			
TITLE	The WashU-HMI Mouse EST Project			
JOURNAL	Unpublished (1996)			
COMMENT	Contact: Marra M/Mouse EST Project WashU-HMI Mouse EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@watson.wustl.edu			



LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	ORGANISM	REFERENCE	AUTHORS	TITLE	COMMENT
AI324867/c	AI324867	546 bp	mRNA	linear	EST 23-DEC-1998				
	mi2b01.1 Soares mouse p3MNF19.5 Mus musculus cDNA clone								
	IMAGE:473065 3' similar to TR:Q28075 Q28075 CHYMOSIN C.; mRNA sequence.								
	AI324867								
	AI324867.1 GI:4059296								
	EST.								
	house mouse.								
	Mus musculus								
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;								
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.								
	1 (bases 1 to 546)								
	Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,								
	Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,U., Morris,M.,								
	Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,								
	Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and								
	Waterson,R.								
	The WashU-HM1 Mouse EST Project								
	Unpublished (1996)								
	Contact: Marra M/Mouse EST Project								
	WashU-HM1 Mouse EST Project								
	Washington University School of Medicine								
	4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108								
	Tel: 314 286 1800								
	Fax: 314 286 1810								
	Email: mouseest@watson.wustl.edu								
	This clone is available royalty-free through LNL; contact the								
	IMAGE Consortium (info@image.llnl.gov) for further information.								
	MG1:263809								
	This clone was previously sequenced on the 5' end only, this new								
	data is from the 3' end								
	High quality sequence stop: 451.								
FEATURES									
source									
	1..546								
	/organism="Mus musculus"								
	/db_xref="taxon:10090"								
	/clone="IMAGE:473065"								
	/clone_lib="Soares mouse p3MNF19.5"								
	/dev_stage="19.5 dpc total fetus"								
	/lab_host="DH10B (ampicillin resistant)"								
	/note="Vector: pTV3D (Pharmacia) with a modified								
	polylinker. Site.1: Not I; Site.2: Eco RI; 1st strand cDNA								
	was primed with a Not I - oligo(dT) primer [5',								
	TGTTACCAATCTGAGTGGAGCGCCGCAATTTTATTTTATTTT 3'],								
	double stranded cDNA was size selected, ligated to Eco RI								
	adapters (Pharmacia), digested with Not I and cloned into								
	the Not I and Eco RI sites of a modified pTV3 vector								
	(Pharmacia). Library went through one round of								
	normalization to a Cot = 5. Library constructed by Bento								
	Soares and M.Fatima Bonaldo. RNA was kindly provided by								
	Dr. Minoru Ku (Wayne State University)."								
BASE COUNT	134 a 140 c 140 g 132 t								
ORIGIN									
Query Match	6.2%; Score 246; DB 9; Length 546;								
Best Local Similarity	70.9%; Pred. No. 1.5e-22;								
Matches 343: Conservative	0; Mismatches 135; Indels 6; Gaps 141								

Oy	2423	cggagcttcagcgacgacatctccacatctcagcagcatctcttgagccacacgaacacgata	2482
Db	366	ggggcctctgtagagacatcttccatattttacacaaagctctttggagctgtgcacaaagccatata	307
Oy	2483	cggtagctttgacatagatctgcagacaaccttagctacatctgacctgaagttgtctttagat	2542
Db	306	tgaccagattttgacatgcactgctgtgagacgttgagacatcattgccaccacgctttgctttgagat	247
Oy	2543	caagcgacaagatgtaccacactggaacccctccgctctataccacgcaagatcaaggtctctg	2602
Db	246	ccatgtgtaggagagttccacacgtccacacctcttagctttatnccaaacaggtacgggctctctg	187
Oy	2603	caccctgtgattccagagtgaggaacacctcccccagaagaatggaatctcttgagagatgtgtcat	2662
Db	186	ctccacgtgcttc-----aaacagggctctccacacatgtggatctcttggggatgttttcatt	133
Oy	2663	tcgttagtactacagcgctcttttgaaagggccaaacacctgttgctgagctaaagcaat	2722
Db	132	tcggagatttatnacaagtgtgttgacagacccaacacatgctgtggctgagccaaagcgtat	73
Oy	2723	ctga	2726
Db	72	ctga	69
RESULT 11			
A1385490			
LOCUS	A1385490	517 bp	mRNA
DEFINITION	mjl603.y1 Soares mouse p3NMF19.5 Mus musculus cDNA clone	linear	EST 27-JAN-1999
DESCRIPTION	IMAGE:482524 5' similar to TR:Q28950 Q28950 PREPROCHYMOSIN		
KEYWORDS	PRECURSOR ; mRNA sequence.		
ACCESSION	A1385490		
VERSION	A1385490.1	GI:4198953	
ORGANISM	EST.		
SOURCE	house mouse.		
REFERENCE	Mus musculus		
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Scurionathi; Muridae; Murinae; Mus. 1 (bases 1 to 517)		
TITLE	Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.		
JOURNAL	The WashU-NCI Mouse EST Project 1999		
COMMENT	Unpublished (1999)		
	Contact: Marra M/WashU-NCI Mouse EST Project 1999		
	Washington University School of Medicine		
	4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA		
	Tel: 314 286 1800		
	Fax: 314 286 1810		
	Email: mouseest@wustl.edu		
	This clone is available royalty-free through LNL ; contact the		
	IMAGE Consortium (info@image.llnl.gov) for further information.		
	MGI:293268		
	This read is a RESSEQUENCE of a previously sequenced mouse clone		
	correct orientation)		
	Possible reversed clone: similarity on wrong strand		
	Seq primer: ~40RP from Gibco		
	High quality sequence stop: 442.		
FEATURES			
SOURCE	Location/Qualifiers		
	1..517		
	/organism="Mus musculus"		
	/db_xref="taxon:10090"		
	/clone="IMAGE:482524"		
	/clone.lib="Soares mouse p3NMF19.5"		
	/dev_stage="19.5 dpc total fetus"		
	/lab_host="DH10B (ampicillin resistant)"		
	/note="Vector: pRT30 (Pharmacia) with a modified		
	polylinker. Site_1: Not I - Site_2: Eco RI; 1st strand cDNA		
	was primed with a Not I - Oligo(dT) primer [5'		





CDS	66..1244	/note="data source:SPTR, source key:P20142, evidence:ISS homolog to GASTRICSTIN PRECURSOR (EC 3.4.23.3) (PEPSINOGEN C)
	putative"	
	/codon_start=1	
	/protein_id="PAB25990.1"	
	/db_xref="GI:12843461"	
	/translation="MKNNVVALDCLPLLEALIVPLPKMKSTETKEQGVLDKFLK NHRIDGOKTHFEKFGDYSVLYPERMAWMDSYVEISITGPQPNFLVLEFDGSSNLW SVYCGDEACTTHTRNPNPKSSSTIYTGQGFSLDYGSLGFFGAYDPLRVAQS.IQVP OEPELSENEGPTNFVYAPQFDGIMGLAYPGSLSSGGATLQMLEGALSDPLFVYVJG SOQSSNGGQIYFVGVDENLITGELTWIYVPOEIVTIDELFLGNQASGWCSSGCC GIYDVTGSLIYMPGQVINDLORTIGAGQEGEVFQVSDVSSIPLTFLVINGQOPFL SPSSYIIQEBGSCNVGLESLSLNAESQPLMTLDVFLRSTIYAFVDMGNNRVGLAPSS	
polyA_signal	1369..1374	
	/note="putative"	
polyA_site	1385	
	/note="putative"	
BASE COUNT	281 a 441 c 336 g 327 t	
ORIGIN		

Query Match	6.0%;	Score 236.2;	DB 11;	Length 1385;
Best Local Similarly	53.7%;	Pred. No. 1.9e-21;		
Matches 602; Conservative	0;	Mismatches 483;	Indels 36;	Gaps 4;

QY	1629	gctgagtaaccgcatctctctctcaacaagtaagtcctctcgtaagcgcttgaagaa	1688
Db	111	GCACGCTTGATCAGAGGTCCCTCGAAMAAAATGAGAGATATCCGGAGACCATGAAAGAA	170
QY	1689	catgacctctagaagaagactctctgtcagaacaacagatgcatcagaagaagtaacc	1748
Db	171	CAAGGTGTCCTCMAAGCACTTCTGGAAGAACCAAGATATGACCTCGGCAGAAATACAC	230
QY	1749	gacctgggtgaagttgctagcgtgcacactacaa-----tacctgatatgtcaa	1799
Db	231	TTTGGCAAGTTTGGTGTGACTACAGTGTACTGTATGACCCCATGGCTTATATGATGATCTTCC	290
QY	1800	tacttggagaagatctactcctcggaacacccgcctcaagaagttacagttctcttatact	1859
Db	291	TACTATGTGTGATCAGACATCGGGACTCCACCCAGCAACTCTCTGTCTTTTGACACT	350
QY	1860	ggtctcctctgaactcttgggttccctctactctactgtcagaagaagatgctctcagaagcaac	1919
Db	351	GGCTCTTCACCAACTGTGGGTGTCTTGTGTCTACTGTGCACAGGAGAGGCTGTGCACCAACAC	410
QY	1920	caagaattgatccgagaagaatcgctcaccttcacagaactttagcgaacacctgtgtcta	1979
Db	411	ACCGGCTACAAACCCACAGATGCTCTCACCCTACTACACTCAAGGACAGACTTTCCTCG	470
QY	1980	caactcgttaacagtlagcatgtaagaagaatcttagctatgataccgttactgtctcacaac	2039
Db	471	CAGTACGGGACACGGCAGACCTTACCGGCTCTTTCGGCTATGACACTCTGAGATCCCAAGAC	530
QY	2040	atgtgacacatcaacagacagaagtagaacttagaacccaagaacacagtgatgtctcaac	2099
Db	531	ATTCAGGTTCCCTAACACGAGAGTTCCGGCTGTAGTGAAGATGAGACCTTGACACCAATTTGTGC	590
QY	2100	tatgcagaatctcgaatgcatccttgttatagtcaataccatcgctcgtcgtcagaagtaactcg	2159
Db	591	TACGCCCAATTGACGGGATCATGTGGGCTTGACCTCCCGGCTGTCTCAAGGGGGGCC	650
QY	2160	ataacctgtttgacaacaatgatagtgaacacgaacactctagtctccaagaacttgtctcggt	2219
Db	651	ACGACGCCCTTTCGACGGGCAATGTTGGGGAGGGGCTCTCTCCACAGCCCTCTTCGGTGTGC	710
QY	2220	tacatgacagaagaatgacgaagagacat-----gtcaacctgtgaagctatgtatcca	2273
Db	711	TACCTTGGCAGCAGGAGGGGTCTAACGCGGGCACAATTGTCTTGCGTGGCTGGACGAG	770
QY	2274	tcctactacaagaatctctctcaacttggttccaagtaactgtgcagcagtaactgtgcaattc	2333

[illegible]

RESULT 13  
AW868692

LOCUS	AW6868692	619 bp	mRNA	linear	EST 22-MAY-2000
DEFINITION	MR1-SNO006-150500-004-a03	SNO006	Homo sapiens	CDNA,	mRNA sequence.
ACCESSION	AW6868692				
VERSION	AW6868692.1	GI:8002744			
KEYWORDS	EST.				
SOURCE	human.				

REFERENCE  
AUTHORS

**AUTHORS**  
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M. R., Nagai, M. A., da Silva, W. Jr., Zago, M. A., Bordin, S., Costa, F. P., Goldman, G. H., Carvalho, A. F., Matsumura, A., Bala, G. S., Simpson, D. H., Brunstein, A., deoliveira, P. S., Bucher, P., Jongeneel, C. V., O'Hare, M. J., Soares, F., Brentani, R. R., Reis, L. F., de Souza, S. J. and Simpson, A. J.

JOURNAL  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
MEDLINE  
20202663  
COMMENT  
Contact: Simpson A.J.G.

Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil

Tel.: +55-11-2704922  
Fax: +55-11-2707001  
Email: [asimpson@ludwig.org.br](mailto:asimpson@ludwig.org.br)  
This sequence was derived from the PAPESP/ICR Human Cancer Genome project. This entry can be seen in the following URL (<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2-mr1-SND064-150>)  
500-004-a03at5-2000-05-15&ta=1)  
Seq primer: puc 18 forward  
High quality sequence start: 10  
High quality sequence stop: 619.

FEATURES  
source

```
/organism="Homo sapiens"  
/db_xref="taxon:9606"
```



```

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="MGD:MGI:1895666"
/db_xref="taxon:10090"
/clone="2210410L06"
/sex="male"
/tissue_type="stomach"
/clone_lib="Riken full-length enriched mouse cDNA library"
/dev_stage="adult"
66..1256
/note="data source:SPRR, source key:P20142, evidence:ISS
homolog to GASTRICIN PRECURSOR (EC 3.4.23.3) (PERSINOGEN
C)
putative"
/codon_start=-1
/protein_id="BAB25952.1"
/db_xref="GI:12843350"
/translation="MKMYVALLCPLLEALIRVPKKKSIETMKEGVAKDLK
MKYDQKTHKFKGSDYSLVPPAAMDASYGEISIGTPPNFLVLEPTGSSNLMY
SSVYQSEACTHTRYNPSKSTYYTQGFSLQYGTGSLTFEGIDTLKVS1QVFN
OERGLSENEPNTNRYAOPDGLMGLAYPGLSSGATTAALQMLBEGALSOPLRVLG
SOOGNGGOIVFGVDENLYTGLTWIPYQELWQITIDFLIGNQASGSSGCO
GIYDGTSLILVMAPOYLINELIOTIGAQEEGYGVFSCDSVSLPILTFVINGVOPFL
SPSSTIIOEBGSCMGLSELISLNAESGQPLMIGDVFILRSYVASTMAITGMALPLIS
RDI"
polyA_signal
1368..1373
/note="putative"
polyA_site
1388
/note="putative"

```

```

BASE COUNT      282 a      441 c      337 g      328 t
ORIGIN

```

```

Query Match      5.7%; Score 227.4; DB 11; Length 1388;
Best Local Similarity 53.9%; Pred. No. 2,6e-20;
Matches 546; Conservative 0; Mismatches 446; Indels 21; Gaps 3;

QY 1629 gctgagatcacccgcatctctctacaaagtaagtcctcgtlaagcgtcgaagaa 1668
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 111 GCACCTTTGATCAGGTCGCCCGCAAGAAATGAAGATATCCGGAGACCATGAAGAA 170

QY 1689 catggaactcagaagactcttcgagaagaacaacatgacacgaagaagatctcc 1748
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 171 CAAGGTGCTCTCAAGACTTTCTGAAGAACCAAGATATGACCTGCGCAGAAATACAC 230

QY 1749 ggtctcggtgaagtcgtaagcgtgccacttaccaa-----ctacctgatagtcaa 1799
    | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 231 TTTCGCAAGTTGGTACTACAGTACTCTATAGACCCCATGGCTTATATGATGCTTCC 290

QY 1800 taacttggaagactctacactcgaaccccgccctcaagaagtcaccggtctcttgatact 1859
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 291 TACTATGGTGATCAGATCGGAGCTCCACCCAGAACTTCTGTCCTTTTGACACT 350

QY 1860 ggttcctctgaactcttggttccctctatactacgtcaagaacatgctctgaagaacac 1919
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 351 GGCCTCTCAACCTGTGGCTGTCTCTGTACTGCGCAGAGGCTTGACACACACAC 410

QY 1920 caaagatcgatccgagaagtcgccaactcagaacttaggcaaacctctgtctata 1979
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 411 ACCCGCTACAAACCCCAAGATCTCTCACCTACTACACTCACTCAAGGCAACACTTCCCTG 470

QY 1980 cactcgttagacagtagcagcaagaaactctagagctatgatacgcgtcaactgctccaac 2039
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 471 CAGTACGGCACCAGGACCTTACCGGCTTCTTGCGCTATGACACTGTGAGATGCCAAGC 530

QY 2040 atgttgacatcaacagacagtaggacttagcaccacaagaacaggtgagtcttcacc 2099
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 531 ATCCAGTCCCTTAACCAAGAGATTGCGCTGAGTGAAGATGAGCCTGCGCACCACAAATTTGTC 590

QY 2100 tatcgaatcagatgacatccttgtagtataccatcgctcgctcagagtaactcg 2159
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 591 TAGCCCAATTGACAGGATCATGGGCTGGCCTTACCCGGCCTGTCTTTCAGGGGGCGCC 650

```

```

QY 2160 ataccgtgtttgacaacatgataaccgaacctagtagtaactgaagactgttctcgtgt 2219
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 651 AACACCCCTTTCAGAGGCACTGTTGGGGAGGCGCTGTCCTGTCACCCCTTCTCGGTGTC 710

QY 2220 tacatggacag-----gaatgcccagagagacatgtctcaacgctttagagctatgatcca 2273
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 711 TACTTTGGCAGCCAGAGGAGGCTTAACGGCGGAGATGTGTGTCGTGGCGGTGAGACAG 770

QY 2274 tctactacacagatctctcactcagtggttccagtcacgtctgcaagactgagcaatc 2333
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 771 AACCTGACACTGGCAGCTACCTGATTCCTGATTCACCCAGAGAGCTTACTGCGACATC 830

QY 2334 actgtgacacagtgacacatcagcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2387
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 831 ACCATTTGACGACTTCTTATTTGGCAACCCAGGCTCTGCTGCTCTCTCTCTGCTGCTG 890

QY 2388 caaactcactctgatacagtagtcccaactgtgtcagactcagacagacatctcaac 2447
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 891 CAAGGCAATTGTAGACACAGGACACTCTCTGCTCATGCTGCGCCAGTACCTGAATGAA 950

QY 2448 attcaagaactatcttgagccacacagaaccagtagcaggttgacatagatcgcgac 2507
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 951 CTTCGACAGACATAGAGAGCCCGAGAGAGATGAGACAGTATTTGTGACCTGCGAC 1010

QY 2508 aaacttagctatacgtcctacagtgcttcttgagatacaaggaagatgataccactgacc 2567
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1011 AGCCTCAGTACCTCCTTACCTCCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCA 1070

QY 2568 cctccgcctctacacagacagatcaaggtgtgtcgaaccagtgatccagag 2620
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1071 CCTCTTCTCTACATCATCCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1123

```

```

RESULT 15
AI322423      414 bp      mRNA      linear      EST 23-DEC-1998
LOCUS
DEFINITION m182B01.y1 Soares mouse p3NMP19.5 Mus musculus cDNA clone
IMAGE:473065 5' similar to TR:Q28075 Q28075 CHYMOSIN C.; mRNA
sequence.
ACCESSION AI322423
VERSION AI322423.1 GI:4056852
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
REFERENCE 1. (bases 1 to 414)
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisler,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellendberg,K., Stepien,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
The WashU-HMI Mouse EST Project
Unpublished (1996)
CONTACT: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LBNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:283809
This read is a RESIDUENCE of a previously sequenced mouse clone
correct orientation)
putative full length read
vector to vector length is 655
Seq primer: -40RP from Gibco
High quality sequence stop: 395.
Location/Qualifiers
I..414

```

FEATURES



**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 21:14:30 ; Search time 140.83 Seconds  
(Without alignments)  
6901.738 Million cell updates/sec

Title: US-09-643-755B-3

Perfect score: 3957

Sequence: 1 cgcagcagatcatctgact.....acctaccacaaagtgacc 3957

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/2/1na/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/2/1na/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/2/1na/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/2/1na/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/2/1na/PTUS\_COMB.seq:\*  
6: /cgn2\_6/ptodata/2/1na/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	897.4	22.7	1240	1 US-08-240-372-2	Sequence 2, Appl
2	895.8	22.6	2733	2 US-08-846-021A-6	Sequence 6, Appl
3	884.6	22.4	2732	6 5217891-14	Patent No. 5217891
4	203.6	5.1	832	1 US-08-318-193-15	Patent No. 5200327
5	203	5.1	828	6 5200327-6	Patent No. 5200327
6	199.8	5.0	846	6 5200327-5	Patent No. 5200327
7	199.4	5.0	838	1 US-08-318-193-13	Sequence 13, Appl
8	192.2	4.9	2073	4 US-09-032-523-6	Sequence 6, Appl
9	146.6	3.7	1299	1 US-08-723-938-4	Sequence 4, Appl
10	146.6	3.7	1299	2 US-09-080-538-4	Sequence 4, Appl
11	146.6	3.7	1910	4 US-08-974-691-7	Sequence 7, Appl
12	146.6	3.7	2061	4 US-09-008-271A-16	Sequence 16, Appl
13	138.2	3.5	2038	2 US-08-631-097-7	Sequence 7, Appl
14	138.2	3.5	2038	4 US-08-810-712-11	Sequence 11, Appl
15	138.2	3.5	2038	4 US-08-810-712-11	Sequence 11, Appl
16	128.6	3.2	1353	4 US-08-974-691-1	Sequence 1, Appl
17	111.4	2.8	1615	1 US-09-640-305-5	Sequence 5, Appl
18	111.4	2.8	1615	1 US-08-360-673-5	Sequence 5, Appl
19	109	2.8	2875	1 US-08-328-314-1	Sequence 1, Appl
20	109	2.8	2875	1 US-08-731-045-1	Sequence 1, Appl
21	97.8	2.5	2032	1 US-08-088-633-1	Sequence 1, Appl
22	97.8	2.5	2032	1 US-08-245-756-1	Sequence 1, Appl
23	97.8	2.5	2032	1 US-08-441-750-1	Sequence 1, Appl
24	97.8	2.5	2032	1 US-08-441-751-1	Sequence 1, Appl
25	97.8	2.5	2032	5 PCT-US92-02521-1	Sequence 1, Appl
26	95.8	2.4	2454	3 US-09-079-415-1	Sequence 1, Appl
27	86	2.2	7218	1 US-08-232-463-14	Sequence 14, Appl

C 28	75.8	1.9	19124	2 US-08-487-826B-13	Sequence 13, Appl
C 29	75	1.9	4810	3 US-08-852-629-11	Sequence 11, Appl
C 30	75	1.9	4838	1 US-08-852-629-15	Sequence 15, Appl
C 31	73.8	1.9	1511	1 US-07-991-867B-8	Sequence 8, Appl
C 32	73.8	1.9	1511	1 US-08-107-755A-8	Sequence 8, Appl
C 33	73.8	1.9	1511	2 US-08-544-332-8	Sequence 8, Appl
C 34	73.6	1.9	1431	4 US-07-991-867B-32	Sequence 32, Appl
C 35	71.4	1.8	660	1 US-09-316-083-2	Sequence 2, Appl
C 36	71.4	1.8	660	1 US-08-107-755A-32	Sequence 32, Appl
C 37	71.4	1.8	660	2 US-08-544-332-32	Sequence 32, Appl
C 38	71	1.8	3095	6 5231168-1	Patent No. 5231168
C 39	69	1.7	615	4 US-08-998-416-186	Sequence 186, App
C 40	68.2	1.7	8920	2 US-08-446-855A-1	Sequence 1, Appl
C 41	68.2	1.7	8920	4 US-09-150-741-1	Sequence 1, Appl
C 42	67.8	1.7	51952	3 US-08-947-823-1	Sequence 1, Appl
C 43	67	1.7	19124	2 US-08-487-826B-13	Sequence 13, Appl
C 44	66.8	1.7	6124	4 US-08-213-419B-3	Sequence 3, Appl
C 45	66	1.7	837	4 US-08-998-416-288	Sequence 288, App

## ALIGNMENTS

RESULT 1

US-08-240-372-2

Sequence 2, Application US/08240372

Patent No. 5711665

GENERAL INFORMATION:

APPLICANT: KATO, ELIE K.

APPLICANT: STUART, W. DORSEY

TITLE OF INVENTION: LIGHT-REGULATED PROMOTERS FOR PRODUCTION

OF HETEROLOGOUS PROTEINS IN FILAMENTOUS FUNGI

NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:

ADDRESS: MORRISON & FOENSTER

STREET: 2000 Pennsylvania Ave. N.W.

CITY: Washington, D.C.

COUNTRY: USA

ZIP: 20006-1812

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/240.372

FILING DATE: 10-MAY-1994

ATTORNEY/AGENT INFORMATION:

NAME: MURASHIGE, KATE H.

REGISTRATION NUMBER: 29,959

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 887-1500

TELEFAX: (202) 887-0763

TELEX: 90-4030

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 1240 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: Linear

US-08-240-372-2

Query Match

Best Local Similarity 87.4%; Score 897.4; DB 1; Length 1240;

Matches 982; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

Oy 1609 tcgttcgttactcagcgtcgtgagatccagcgcattctctctacaagtgatctc 1668

Db 100 TCTTCGCTCTCTCCAGGCGCTGAGATCCAGATCCCTCTGTACAAAGCAAGTCTC 159





Db 1930 GACCCGAGAAAGTCCTCACCCTTCAGAACCTGGGCAAGCCCTCTCTATCCACTACGGG 1989  
Oy 1989 acaagtagcaagaagaatcttagctatgataaccgtacgtctccaaactgtgac 2048  
Db 1990 ACAGGACACATGACGAGGATCTGGGCTATGACACCGTACGTACTCTCCAACTTTGGGAC 2049  
Oy 2049 attcaacagaagtagaacttagcaaccaagaaccagtgatgtcttcacatgcaga 2108  
Db 2050 ATCCAGGACAGAGTAGGCGCTGAGCAACCGAGAGCCGGGAGCTCTTCCATCCATGCCGA 2109  
Oy 2109 ttcatgtgcattccctgtgtatgacataccatcgctcgctgaagtaactcgatctgt 2168  
Db 2110 TTCCGAGGAGATCCTGGGATGAGCTTACCCCTTCGCTCAGAGTACTCGATACCCGCTG 2169  
Oy 2169 tttagcaacatgataagcagcaactagtagtaagaactttctcggtttacatgac 2228  
Db 2170 TTTGACACATGATATMAAGGCACTGGTGGCCAGACCTTCTCGGTTTACATGGAC 2229  
Oy 2229 aggaatggccaggaagagatgtctcaagcttggagctatgtatccatctactacaaga 2288  
Db 2230 AGGATGGCCAGGAGAGATGCTACGCTGGGGCCATCGACCCGTCCTACTACACAGGG 2289  
Oy 2289 tctctcaactgtgtccagtaactgtgtgacagactctggcaattcaactgtgacagtgctc 2348  
Db 2290 TCCCTGCACTGGGTGCTCCGTCACACTGACAGCAGTACTGGCACTGCTGAGACAGTGTc 2349  
Oy 2349 accatcaacggt 2408  
Db 2350 ACCATCAGCGGT 2409  
Oy 2409 acgtccaaagt 2468  
Db 2410 ACCTCAAGCTGT 2469  
Oy 2469 acaagaacacgtacgt 2528  
Db 2470 ACACGAACCGATGAGGT 2529  
Oy 2529 gt 2588  
Db 2530 GTGGCTTTTGATATMAAGGCAAAATGTACCCACTGACCCCTCCGCTATACCCAGCCA 2589  
Oy 2589 gatcaaggt 2648  
Db 2590 GACCAAGGCTTCTGTACCAAGTGGCTTCCAGATGAAATCATTCACAGAAATGATCCCG 2649  
Oy 2649 ggaagatgt 2708  
Db 2650 GGGGATGTTTTCATCCGAGAGTATACAGCGCTTGTGACAGGGCAACACTCGTGGGG 2709  
Oy 2709 ctgagtaagaatctgaa 2727  
Db 2710 CTGGCCAAAGCCATCTGAA 2728

RESULT 3  
5217891-14  
: Patent No. 5217891  
: APPLICANT: BRAKE, ANTHONY J.;VAN DEN BERG, JOHAN A.  
: TITLE OF INVENTION: DNA CONSTRUCTS CONTAINING A KUDYEROMYCES  
: A FACTOR LEADER SEQUENCE FOR DIRECTING SECRETION OF HETEROLOGOUS  
: POLYPEPTIDES  
: NUMBER OF SEQUENCES: 23  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/507,398  
: FILING DATE: 09-APR-1990  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: 78,551  
: FILING DATE: 28-JUL-1987  
: SEQ ID NO:14:  
: LENGTH: 2732  
5217891-14

Query Match 22.4%; Score 884.6; DB 6; Length 2732;  
Best Local Similarity 82.6%; Pred. No. 6,9e-173;  
Matches 1024; Conservative 0; Mismatches 214; Indels 1; Gaps 1;  
Oy 1610 cgttctgttactacagcgt 1669  
Db 669 cgaatgttccatcatatgtcgtgagatcacagatctccctctgtacaagaagctctct 728  
Oy 1670 ccgttaagcgtctgaaggaacatgagctctagaagaactcttgcagaanaacagatag 1729  
Db 729 gaggaaagcgcgtgaaggaacatgagctcttgcagaagaactcttgcagaanaacagatag 788  
Oy 1730 catcagaacgaagtactccggtctcggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1789  
Db 789 catcagaacgaagtactccggtctcggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 848  
Oy 1790 tgatgtcaataacttggggaagatctactctcgaaaccccgctcaagaatctacgttct 1849  
Db 849 ggacagctcaagtacttgggaagatctactctcgaaaccccgctcaagaatctacgttct 908  
Oy 1850 cttgtatctgttctctcgtacttctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1909  
Db 909 gtttgaactgtgtctctgtacttctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 968  
Oy 1910 caagaacacacaaagt 1969  
Db 969 caaaacacacacagcgtctgt 1028  
Oy 1970 ctgtctatatactacgt 2029  
Db 1029 cctgtctatatactacgt 1088  
Oy 2030 tgtcttccaacttgt 2089  
Db 1089 tgtcttccaacttgt 1148  
Oy 2090 tgtcttccaacttgt 2149  
Db 1149 cgtcttccaacttgt 1208  
Oy 2150 agatgtactgt 2209  
Db 1209 agatgtactgt 1268  
Oy 2210 gtctcgtgttcatgt 2268  
Db 1269 gtctcgtgttcatgt 1328  
Oy 2269 atccatctactaacaagaatctcttcaactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2328  
Db 1329 acccgttccactaacaagaatctcttcaactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1388  
Oy 2329 aatttactgt 2388  
Db 1389 agtttaactgt 1448  
Oy 2389 aagctacttgt 2448  
Db 1449 aggcatacttgaacacggt 1508  
Oy 2449 ttcaagaacgttgaacacggt 2508  
Db 1509 ttcaagaacgttgaacacggt 1568  
Oy 2509 accttgaactacgt 2568  
Db 1569 accttgaactacgt 1628  
Oy 2569 cctccgctataccagcagaagatcaagggttctgcacagtgatctccagatgtgaagacc 2628  
Db 1629 cctccgctataccagcagaagatcaagggttctgcacagtgatctccagatgtgaagatc 1688





```

STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/032,523
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0479 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 2073 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: PANTUM01
CLONE: 151565
US-09-032-523-6

```

```

Query Match          4.9%: Score 192.2; DB 4; Length 2073;
Best Local Similarity 56.5%: Pred. No. 8.3e-33;
Matches 406; Conservative 0; Mismatches 296; Indels 15; Gaps 2;

```

```

OY 1640 ccgacttcctctcaagaagtaagtcctccgtaagcgctgaaggaacatgacttct 1699
    |||||
DB 104 CAGGTTGCCCTCAGAGAGGATCCGTCCTCAAGAAAGCTGGCGGACGAGCAGCT 163
OY 1700 agaagactcttcgaaacaacaglatgcaatca-----gcagaagctactccgctt 1753
    |||||
DB 164 CTCGTGATTTCTGGAATCCATTAATTTGGACATGCCAGTTCACCGAGTCTGTCAT 223
OY 1754 cgtgaagtgtgtagcggtgcacactcaactcattagatcaactattgggaagat 1813
    |||||
DB 224 GACCCAGAGTGCAGAGAACCCCTCATCAACTGATGGATGATGATGATCTTGGCAGTAT 283
OY 1814 ctaactcgaaaccgacctcaagagttcacgctctctcttgaactgttccctgaact 1873
    |||||
DB 284 CTCCTATTGGCTCCACACACAGAACTTCACTCTCATCTTCGACACTGGCTCTCAACT 343
OY 1874 ctgggttcctctatctactgaagaaatgcctgcaagaacacccaagaattcgatcc 1933
    |||||
DB 344 CTGGGTCCTCTGTGTACTGTGACTGACAGCCAGCTGCAAGACGACAGAGGTTCCAGCC 403
OY 1934 gagaagaatcgtcacacttcagaacttaggcaaacctgtctatatacatagttacag 1993
    |||||
DB 404 TTCCAGTCCAGCATATACAGCCAGGAGTCAATCTTCTCATTTCACTATGGAACCG 463
OY 1994 tagcatgaaggaaattcaggctatgataccgtcaactgtctcaaatcgtgacattca 2053
    |||||
DB 464 GAGCTTTGTCGGGATCATTTGAGACCGACCAAGTCTCTGTGGAAGAGCTAACCGTGTG 523
OY 2054 acagacagtaggaacttagcaccacaagaacaggtgtagtcttcaactatgcagaattcga 2113
    |||||
DB 524 CCAGCAGTTTGGAGAAAGGTGCACAGAGCCAGGCGCAGACCTTTGTGATGCAAGATTGA 583
OY 2114 tggactcttgtagatgatacccaatcgctcggtgtagagtagtactatcctggtgttga 2173
    |||||

```

```

DB 584 TGAATTCCTGGGCTGGGATACCCCTTCTTGCTGTGGAGAGAGTGACTCCAGTATTGA 643
OY 2174 caaatgatgaaccgacacttagtagtcaagacttgctcgtttacatgacagaa 2233
    |||||
DB 644 CAACATGATGGCTCAGAACCTGGTGACTTGGCCGATGTTTCTCTCATGAGACAGTAA 703
OY 2234 -----tggcagagagacatgctcagcgttgagactatgatalcactactaac 2284
    |||||
DB 704 CCCAGAAAGTGTCGCCGGAGCGAGCTGATTTTGGAGGGCTACGACCACTCCATTCTC 763
OY 2285 agaatcttcaactggtgtccagtcactgtgcagcagtagtgcgaattcactgtgaca 2343
    |||||
DB 764 TGGAGCCTGAATTGGTCCAGTACCAACAGACTTACTGGAGATGGACTGTGATA 822

```

```

RESULT 9
US-08-723-938-4
Sequence 4, Application US/08723938
Patent No. 5776759
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Coleman, Roger
TITLE OF INVENTION: TWO NOVEL HUMAN CATHESPIN PROTEINS
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESS: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: U.S.
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/723,938
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0125 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1299 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ANTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE:
ORIGINAL SOURCE:
IMMEDIATE SOURCE:
LIBRARY: LUNGNOT02
CLONE: 312099
US-08-723-938-4
Query Match          3.7%: Score 146.6; DB 1; Length 1299;
Best Local Similarity 50.4%: Pred. No. 1.7e-21;
Matches 421; Conservative 0; Mismatches 399; Indels 15; Gaps 2;
OY 1769 cgtgcacttacaactacttgatgataactcttggaagatctactcgaacccc 1828

```

```

Db 217 CGTACCTCTCGAACTACAGAGATGTCAGATATTTGGGAAATTTGGGGAACCC 276
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1829 gccccaagatcaccgtctccttgatactggtccctcctgactcgtggtccctccta- 1887
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 277 TCCACAAAACCTTCACTGTGGCTTGGACATGGCTCCACATCTCTGGGTCCCGCAG 336
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1888 -----tctactgcaagcaatgctgcgaagaaccaccaagattcgatccgaagaagtc 1942
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 337 GAGATGCCACTCTTTCAGTGTGCCCTGCTGGTTACACACCGATTGTATCCAAAGCCTC 396
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1943 gtccacctccagaagacttaggcaaaccttgctctatacactaggttaaggtatgca 2002
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 397 TAGCTCCCTTCCAGGCCAATGGAGCAAGTTGGCCATTCATATATGAACTGGCGGCTAGA 456
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 2003 agaatcttagctatgatactgctacgtctcctcgaacattgtgacattcaacagaagtc 2062
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 457 TGGAAATCCTGAGCGAAGCAAGCTGACTATTTGGGAATCAAGGGTGCAATGATATTTT 516
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 2063 aggaactagaccaccaagacaggtgatgtcttcacctatgcaagattcgatgcatcct 2122
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 517 CGGGAGGCTCTCTGGAGGCCACCGCTGTCTTCCCTTTTGGCCATTTTGAATGGATATTT 576
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 2123 tggatagcataccatcgctcgctgcagagtaactgatacctgtgttgacaacatgat 2182
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 577 GGGCCTCGGTTTCCCATTTCTGTGTGAAGAGATTGGCCCGCATGATGATGATGAT 636
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 2183 gaaccgacaacctagtagtcaagactgtgtctggtttacatgagcaagaatgagcaaga 2242
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 637 GGAGCAGGGGCTATTTGATTAAGCTGTCTTCTTCTTTTACCTCAACAGGAGCCTGAAGA 696
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 2243 g-----agcatgctcaagcttgagactatgatcatcctactacacagaagtcct 2293
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 697 GCCGTGATGAGAGAGAGCTGGTTCCTGGGGGCTCGGAGCCCGGACACTATCCACCCCT 756
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 2294 tcaactgggttcagctactgtgcaagcagtaactggaactcgtgtgaaagtgtaacat 2353
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 757 CACCTTCTGTCAGTCACAGGTCCCGCCTACTGCGAGATTCACATGAGGAGCGTGAAGT 816
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 2354 caggggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2413
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 817 GGGGCCAAGGCTGACTCTGTGTGCCAAGGGCTGTGTCTCCATCTTGATGACGGGAGCTC 876
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 2414 caagctgtgctgagctagcaagacatctcaacatcagcaagctatgtgagccacaca 2473
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 877 CCTCATCAGAGACCCATGAGAGATCCGGGCCCTGCATGACGACCATTTGGGGGAAATCC 936
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 2474 gaacagtaagtggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2533
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 937 CTTCGTGCTGGGAGTACATCATCTGTGTCGGAATCCCAAGGCTCCCGCAGTCTC 996
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 2534 cttgagagtaaggaagcaagatgtacccactgaacccctcgctctatacagaagcag 2588
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 997 CTTCCTTTTGGGGGGCTGTGTGTAACTCAACGCCCATGATTAATGATATCCAG 1051
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 10  
US-09-080-538-4

```

: Sequence 4, Application US/09080538
: Patent No. 5965129
: GENERAL INFORMATION:
: APPLICANT: Bandman, Olga
: APPLICANT: Coleman, Roger
: TITLE OF INVENTION: TWO NOVEL HUMAN CATHESPIN PROTEINS
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Incyte Pharmaceuticals, Inc.
: STREET: 3174 Porter Drive
: CITY: Palo Alto
: STATE: CA
: COUNTRY: U.S.
: ZIP: 94304
: COMPUTER READABLE FORM:

```

```

: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FASTSEQ Version 1.5
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/080,538
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/723,938
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Billings, Lucy J
: REGISTRATION NUMBER: 36,749
: REFERENCE/DOCKET NUMBER: PF-0125 US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 415-855-0555
: TELEFAX: 415-845-4166
: TELEX:
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1299 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: MOLECULE TYPE: cDNA
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: FRAGMENT TYPE:
: ORIGINAL SOURCE:
: IMMEDIATE SOURCE:
: LIBRARY: LUNGNOT02
: CLONE: 312099
: US-09-080-538-4

```

Query Match 3.7%; Score 146.6; DB 2; Length 1299;  
Best Local Similarity 50.4%; Pred. No. 1.7e-21;  
Matches 421; Conservative 0; Mismatches 399; Indels 15; Gaps 2;

```

Qy 1769 cgtgcacttaccactacactccttgatagtcataactcttggagaatcctcctcgaaaccc 1828
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 217 CGTACCTCTTCGAACTACAGAGATGTCAGATATTTGGGAAATTTGGGCTGGGAACCC 276
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1829 gccccaagatcaccgtctccttgatactggtccctcctgactcgtggtccctccta- 1887
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 277 TCCACAAAACCTTCACTGTGGCTTGGACATGGCTCCACATCTCTGGGTCCCGCAG 336
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1888 -----tctactgcaagcaatgctgcgaagaaccaccaagattcgatccgaagaagtc 1942
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 337 GAGATGCCACTCTTTCAGTGTGCCCTGCTGGTTACACACCGATTGTATCCAAAGCCTC 396
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1943 gtccacctccagaagacttaggcaaaccttgctctatacactaggttaaggtatgca 2002
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 397 TAGCTCCCTTCCAGGCCAATGGAGCAAGTTGGCCATTCATATATGAACTGGCGGCTAGA 456
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 2003 aggaacttaggctatgataccgtacatgctcctcaacatgtggaacattcaacagaagtc 2062
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 457 TGGAAATCCTGAGCGAAGCAAGCTGACTATTTGGGAATCAAGGGTGCAATGATATTTT 516
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 2063 aggaacttagcaaccaccaagacaggtgatgtcttcacctatgcaagattcgatgcatcct 2122
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 517 CGGGAGGCTCTCTGGAGGCCACCGCTGTGTCTTGGCCATTTTGAATGGGATATTT 576
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 2123 tggatagcataccatcgctcgctgcagagtaactgatacctgtgttgacaacatgat 2182
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 577 GGGCCTCGGTTTCCCATTTCTGTGTGAAGAGATTGGCCCGCATGATGATGATGAT 636
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 2183 gaaccgacaacctagtagtcaagactgtgtctggtttacatgagcaagaatgagcaaga 2242
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 637 GGAGCAGGGGCTATTTGATTAAGCTGTCTTCTTCTTTTACCTCAACAGGAGCCTGAAGA 696
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 2243 g-----agcatgctcaagcttgagactatgatccatcctactacacagaagtcct 2293
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```





TOPOLOGY: Linear  
MOLECULE TYPE: Genomic DNA  
HYPOTHEetical: No  
ANTI-SENSE: No  
FRAGMENT TYPE: No. 5968816 applicable  
ORIGINAL SOURCE:  
ORGANISM: homo sapiens  
STRAIN: not applicable  
INDIVIDUAL ISOLATE: not applicable  
DEVELOPMENTAL STAGE: not applicable  
HAPLOTYPE: not applicable  
TISSUE TYPE: blood  
CELL TYPE: Leucocyte  
ORGANELLE: not applicable  
IMMEDIATE SOURCE:  
LIBRARY: not applicable  
CLONE: not applicable  
POSITION IN GENOME:  
CHROMOSOME/SEGMENT: not applicable  
MAP POSITION: not applicable  
UNITS: not applicable  
FEATURE:  
NAME/KEY: This is the DNA sequence  
NAME/KEY: claimed in 15(v1) as the Cathepsin gene in FIG. 15.  
LOCATION: not available  
IDENTIFICATION METHOD: experiment-  
IDENTIFICATION METHOD: in specification  
OTHER INFORMATION: prevention of IFN-2  
OTHER INFORMATION: promoted cell death  
PUBLICATION INFORMATION: not available  
US-08-631-097-7

Query Match 3.5%; Score 138.2; DB 2; Length 2038;  
Best Local Similarity 51.0%; Pred. No. 1e-19; Mismatches 378; Indels 48; Gaps 3;  
Matches 443; Conservative 0;

QY 1769 cgtgcacttaccacttacttgatagtcataacttgggaagatctactcggaaacccc 1828  
DB 255 CGAGGTCCTCAAGAACTACATGAGCCCGCAGTACTCGGGGAGATGGCATGGGAGCGC 314  
QY 1829 gctcaagaagttcaaccgtctcttggatagctggttcctctgacttgggttcctctat 1888  
DB 315 CCCCCAGTCTTCAAGTGTCTTCGACACGGGCTCTCAACCTGTGGTCCCTCCAT 374  
QY 1889 ctactgcaaa-----gagcaatgcttgcgaagaacacccaagaatctgacccgaaagtc 1942  
DB 375 CCACCTGCAAACTGCTGGACATGCTTGTGATCCACCAAGTACAAACGACAGATC 434  
QY 1943 gtccaccttccagaacttaggcaaaccttgcctatacactacggtacaggtagcatgca 2002  
DB 435 CAGCACTCACTGTAAGAAAGTACCTCGTTTGACATCCACTATGCTCGGGGACCTCTC 494  
QY 2003 aggaatcttaggctatgataccgttcaactgtctccaaacttgggaattcaa----- 2054  
DB 495 CGGTAACCTGAGCAGACACTGTGCGGTCGCCCTGCCAGTCAAGCTCGTCAAGCTCTGC 554  
QY 2055 -----cagacagtaggaacttagcaacccaagaacaggtga 2089  
DB 555 CCTGGGCGGTGCAAAAGTGGAGAGGCTCTTTGGGGAGGCGCCACCAAGCAGCAGGCAT 614  
QY 2090 tctcttcaactatgcagaatctgatgacatccttggtatggaataccatcgtctgcgtc 2149  
DB 615 CACCTTCATCGCAGCAAGTATGATGCTGGGCGCATGGGCTTACCCCGCATCTCCG 674  
QY 2150 agagtaactgatactgttlttgacaacatgatacgcagacactgtatgcttaagact 2209  
DB 675 CAACAACGTGCTGCCGCTTCGACAACTGATGCAGCAGAACTGTGTGAGCAGAACT 734  
QY 2210 gttctcgtgttcaatgagacagaa-----tgccagagagagatctcaacgtcttg 2260  
DB 735 CTTCCTCTTACTGAGCAGGAGCCAGATGCGCAGCCTGGGGGTGAGTGTGCTGGG 794

QY 2261 agctatgatccatctactacacagagatctcttcaactggttccagtcactgtgcaga 2320  
DB 795 TGGACACAGCTCCAGTATTACAGAGGTTCTGTCTACTGATGATCACCAGCAAGC 854  
QY 2321 gtaactggcaattcactgttggagaagtgtaaccatcagcggtgtgtgttgcatagtgaag 2380  
DB 855 CTATCGCAGGTCCACCTGAGCAGCAAGTGGAGGTGGCCAGCGGCTGTACCTTGAAGGA 914  
QY 2381 tgaatgcaagctatcttggataccggtacggttcaacgttggctggagacttagcaagacat 2440  
DB 915 GGGCTGTAGGCGCATTTGTGACACAGGCACTTCCCTCATGTGTGGGCGCGGTGATGAGGT 974  
QY 2441 tctcaacatlcagcaagctatggaagccacacagaacaaagtaacgttggattgacataga 2500  
DB 975 GCGCGAGCTGCAGAAAGCGCATCGGGCGCGCTGATTCAGGCGAGTACATGATCC 1034  
QY 2501 ttgcgacaaccttagctatactgactacagttgtcttggatgatacagcgaaagtatccc 2560  
DB 1035 CTTGTGAAGAGTGTCCACCCTGCCCGCATCACTGAAGCTGGGAGCAAGGCTACAA 1094  
QY 2561 actgacccccctcgcctataccagcaag 2589  
DB 1095 GCTGTCCCAAGAGACTACACCTCAAG 1123

RESULT 14  
US-08-810-712-11  
Sequence 11, Application US/08810712G  
Patent No. 6160106

GENERAL INFORMATION:  
APPLICANT: Yeda Research and Development Co. LTD  
TITLE OF INVENTION: Tumor Suppressor Genes, Proteins Encoded Thereby and  
TITLE OF INVENTION: Use of said Genes and Proteins  
FILE REFERENCE: sequence list  
CURRENT APPLICATION NUMBER: US/08/810,712G  
CURRENT FILING DATE: 1997-03-03  
EARLIER APPLICATION NUMBER: PCT/US94/11598  
EARLIER FILING DATE: 1994-10-12  
NUMBER OF SEQ ID NOS: 31  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 11  
LENGTH: 2038  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-08-810-712-11

Query Match 3.5%; Score 138.2; DB 4; Length 2038;  
Best Local Similarity 51.0%; Pred. No. 1e-19; Mismatches 378; Indels 48; Gaps 3;  
Matches 443; Conservative 0;

QY 1769 cgtgcacttaccacttacttgatagtcataacttgggaagatctactcggaaacccc 1828  
DB 255 cgagtggttccaagaactatagagcccaagtacacggtacggtgagatggtgagcgc 314  
QY 1829 gctcaagaagttcaaccgtctcttggatagctggttccttgaacttgggttcctctat 1888  
DB 315 ccccgagttcttaacagctgcttgcacaagcgctctccaacacttgggtccctccat 374  
QY 1889 ctactgcaaa-----gagcaatgcttgcgaagaacacccaagaatctgacccgaaagtc 1942  
DB 375 ccacttgcaacttggatagctgttgcgtggaatcccaaggtatacaacagcgaagtc 434  
QY 1943 gtccaccttccagaacttaggcaaaccttgcctatacactacggtacaggtagcatgca 2002  
DB 435 cagcacttcggtgaagaataggtacccgttgcataccactatggtcgtggcagcctcc 494  
QY 2003 aggaatcttaggctatgataccgttcaactgtctccaaacttgggaattcaa----- 2054  
DB 495 cgggtatcccgagcagagacactgttgcgttgccttgcagtcagcgctgcagcctctgc 554  
QY 2055 -----cagacagtaggaacttagcaacccaagaacaggtga 2089



Db	555	cctgsgcggtgtcaagttgagagaggaaggtctctttgsgagccaccagaagcagcgat	614
Qy	2090	tgcttcacctatgacgaaltcgatgcatccttggatgcatatcgcatcgcgcgc	2149
Db	615	caactcatcgacgaagtcgatgcatcctgscatgscatgctaccgccgatctcgt	674
Qy	2150	agagtactcgatcacocigtgtgttgacaacatgatgaaacgcacccatgactcgaagatc	2209
Db	675	caaaagatgctgcgccgtcttcgcacaccttgatgacgacgaagctgtgtgacaaagacat	734
Qy	2210	gtctcgtgttacaatgtagcaggga-----tgccagagagcatgctacgccttgg	2260
Db	735	ctcttcctcttaactgagcagggaaacccagatgcgcagcctgsggtgtgagctgagctggg	794
Qy	2261	agcatattgatcatcactbactacacagagatccttcaacttgggttccagatcactgtgca	2320
Db	795	tggcagacagctccaagatattacaagtgctctgtctgtactacatgacatgacacccgaagc	854
Qy	2321	gtactggtcaattcaactgtgtgacagatgacacatcagcgggtgtgtgtgtgtcatgtgaag	2380
Db	855	ctactgtcgaagttccacacttggaccaggttggaaggltgagcagcgggcgtgacccctgtga	914
Qy	2381	tggatgtcgaagctatccttggatccggtatgcgcgaagctgtgtcgaacttagcagacat	2440
Db	915	gggctgtgagagccatgtgtgacacagagcacttcctcactgctgtggccgggtgagtga	974
Qy	2441	tcctaaatcttcaacagctatgttgagacacacagaaacccagtagctgtgattgacataga	2500
Db	975	ggcgagactgtgcagaagggcatcgtggscgtgtgcctgcatcagggcgtgatacatgatcc	1034
Qy	2501	ttgcgacacacttagctacatgctctacagttgtcttggagatcaacggaagatgtacc	2560
Db	1035	ctgtgaaagaggtgtccacccctgcgcgcatcacacttgaaatcgtgagagcgaagcttaca	1094
Qy	2561	actgacccctccgcactatcacgcagcagg	2589
Db	1095	gtgtcccccagagactaacgcgtcaag	1123
RESULT 15			
PCT-US92-08090-1			
Sequence 1, Application PC/TUS9208090			
GENERAL INFORMATION:			
APPLICANT: Tang, J. N.			
TITLE OF INVENTION: Fusion Protein Genes for Treatment of			
TITLE OF INVENTION: Aids			
NUMBER OF SEQUENCES: 15			
CORRESPONDENCE ADDRESS:			
ADDRESSEE: Kilpatrick & Cody			
STREET: 100 Peachtree Street			
CITY: Atlanta			
STATE: Georgia			
COUNTRY: U.S.			
ZIP: 30303			
COMPUTER READABLE FORM:			
MEDIUM TYPE: Floppy disk			
COMPUTER: IBM PC compatible			
OPERATING SYSTEM: PC-DOS/MS-DOS			
SOFTWARE: PatentIn Release #1.0, Version #1.25			
CURRENT APPLICATION DATA:			
APPLICATION NUMBER: PCT/US92/08090			
FILING DATE: 19920922			
CLASSIFICATION:			
ATTORNEY/AGENT INFORMATION:			
NAME: Padst, Patrea L.			
REGISTRATION NUMBER: 31,284			
REFERENCE/DOCKET NUMBER: OMFI129			
TELECOMMUNICATION INFORMATION:			
TELEPHONE: 404-572-6508			
TELEFAX: 404-572-6555			
INFORMATION FOR SEQ ID NO: 1:			
SEQUENCE CHARACTERISTICS:			

```

1  LENGTH: 2465 base pairs
2  TYPE: NUCLEIC ACID
3  STRANDEDNESS: single
4  TOPOLOGY: linear
5  MOLECULE TYPE: DNA (genomic)
6  HYPOTHETICAL: NO
7  ANTI-SENSE: NO
8  ORIGINAL SOURCE:
9  ORGANISM: Homo sapien
10 TISSUE TYPE: Epithelial
11
12 FEATURE:
13  NAME/KEY: misc_feature
14  LOCATION: 1..6
15  OTHER INFORMATION: /note= "Restriction site"
16
17 FEATURE:
18  NAME/KEY: misc_feature
19  LOCATION: 2460..2465
20  OTHER INFORMATION: /note= "Restriction site"
21
22 COT-US92-08090-1

```

Query Match	3.5%	Score 138.2	DB 5	Length 2465
Best Local Similarity	51.0%	Pred. 1.1e-19		
Matches 443	Conservative 0	Mismatches 378	Indels 48	Gaps 3
QY 1769	ctgtgcacttaccactactaccttgatagtaactcttggagaatcctacccggaacccc	1828		
Db 1424	CGAGGTGCTCAAGAACTCATGATGAGCCGACACTACAGGGGAGATTGGCATGGGAGCGC	1483		
QY 1829	gacctaaagatctaaccttctctcttgatacttggtctctctgaatcttggttccctctat	1888		
Db 1484	CCCCAGTCCTTCACAGTCGTCCTTGGACACGGGGCTCTCCAACTGTGGGTCCCTCCAT	1543		
QY 1889	ctactgcaca-----gaagcatgctctgaagaaccccaagaatctgcacgagaagagc	1942		
Db 1544	CCACTCGCAAACTGCTGGACATGCTGCTTGGCTGATCCACCACTCAAGTACACAGCAGCAAGTC	1603		
QY 1943	gtccaccttcagaacttagcgaacacctgtctatacactcggltacaggtatagcatgta	2002		
Db 1604	CAGCACCTACAGTCAAGAAATGATGTAACCTCGTTTGACATCCACTATGGCTCGGGACCTGTC	1663		
QY 2003	aggaaatcttaggtatgataccgttcaactgctccaaacttggagaattcaaa-----	2054		
Db 1664	CGGGTACTGTGACCCAGGACACTGTGCGTGCCCTGCCACATCAGGCTGCTCAGCCTTGC	1723		
QY 2055	-----cagacaagtagaacttagcaccacaagaacaggtga	2089		
Db 1724	CCTGGCGGTGTCMAAGTGGAGAGGCGAGGCTTTTGGGAGAGGCCACCAACACCGACAGCAT	1783		
QY 2090	tgcttcaactatgcagaattcogatgtagcatccttgatagtacatacccatcgctcgatc	2149		
Db 1784	CACCTTCATCCAGCAGCAAGTTGATGGGATCTCTGGGACATGGGCTTACCCCGCATCTCCGT	1843		
QY 2150	agagttacccgtactcgtgttgagaacatgataggaaacccactagtagtcaagact	2209		
Db 1844	CAACAGACGTGCTGCCCGCTCTTGACACACTGATGACGACGAAAGCTGTGTGACACAGAACT	1903		
QY 2210	gtctcggtttacaatgtagcagaagaa-----tgccagaagagcatgtaacagcttg	2260		
Db 1904	CTTCTCCTTTCACCTGAGCAGCAGGAGCCAGATGCCAGGCTGGGGGTGAGCTATGCTGGG	1963		
QY 2261	agctatgatacactcactactacaacagaatctcttcaactggtgtccagtaactgtgtcaaga	2320		
Db 1964	TGGGACAACATCTCAAGTATTTACCAAGGGTCTCTGTGCTCACTCAAGTATGTCACCCGCAAGGC	2023		
QY 2321	gtactgcgaattcaactgtgagcagatgataccaatcagcgggtgtgtgtgtgtatgtaagag	2380		
Db 2024	CTACTGGCAGGTTCACCTCGACACAGCTGGAGTGGAGTGGCCACGGGCTGACCTTGCAAGA	2083		
QY 2381	tgatgtgtcaagctatcttgatlaaccggtlaagtcacaagctgltcggaaacttagcagcgaat	2440		
Db 2084	GGGCTGTGAGGCTCAATTGTGGACACAGCACTTCCCTCATGTGGTGGGCGGCTGATGAGGT	2143		



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 21:01:50 ; Search time 9238.6 Seconds  
(Without alignments)  
2656.986 Million cell updates/sec

Title: US-09-643-755B-1  
1173  
Perfect score: 1 atgaacttccttaagctctt.....ggctagctaaagcaatctga 1173  
Sequence:

Scoring table:  
GAPOP-60.0-, Gapext 60.0

Searched: 1797656 seqs, 10463268293 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :  
1: gb\_ba:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_inv:\*  
32: em\_htg\_other:\*  
33: em\_htgo\_inv:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
------------	-------	-------------	--------	-------	-------------

1	1173	100.0	1173	6	AX088019	AX088019 Sequence
2	1173	100.0	3957	6	AX088021	AX088021 Sequence
3	56	4.8	1269	4	BOVCHYMOC	J00004 Bovine chym
4	49	4.2	1098	6	E00079	E00079 CDNA sequen
5	49	4.2	1098	6	E02341	E02341 CDNA sequen
6	49	4.2	1101	4	AF177290	AF177290 Bubalus a
7	49	4.2	1115	12	SYNPROCA	M22593 Synthetic b
8	49	4.2	1175	6	I04058	I04058 Sequence 5
9	49	4.2	1240	6	AR002347	AR002347 Sequence
10	49	4.2	1275	4	BOVCHYMOC	J00002 bovine chym
11	49	4.2	1289	6	E00144	E00144 CDNA encodi
12	49	4.2	1291	6	A15836	A15836 chymosin ge
13	49	4.2	1305	4	BOVCHYMOB	J00003 bovine chym
14	49	4.2	1460	6	E00042	E00042 DNA coding
15	49	4.2	1460	6	E00295	E00295 CDNA encodi
16	49	4.2	2726	6	I08097	I08097 Sequence 5
17	49	4.2	2733	6	AR073077	AR073077 Sequence
18	49	4.2	2962	6	I08098	I08098 Sequence 1
19	44	3.8	1094	4	BP019786	U19786 Bos primige
20	38	3.2	1290	6	E00075	E00075 CDNA encodi
21	38	3.2	1291	6	A15633	A15633 preprochymo
22	38	3.2	1311	6	E00108	E00108 DNA coding
23	37	3.2	119	4	BOVCHYMOC	M14076 Bovine chym
24	35	3.0	171	4	BOVCHYMOC	M14070 Bovine chym
25	35	3.0	637	6	E05472	E05472 DNA sequenc
26	33	2.8	179	6	I04059	I04059 Sequence 6
27	33	2.8	1172	4	CDR131677	A131677 Camelus d
28	32	2.7	1335	4	BOVCHYMOC	M14077 Bovine chym
29	32	2.7	1292	4	OAPECHY	X53037 Lamb mRNA f
30	30	2.6	134	4	BOVCHYMOC	M14074 Bovine chym
31	30	2.6	165	4	BOVCHYMOC	M14075 Bovine chym
32	29	2.5	1460	6	E00042	E00042 DNA coding
33	29	2.5	1460	6	E00295	E00295 CDNA encodi
34	27	2.3	139	4	BOVCHYMOC	M14072 Bovine chym
35	25	2.1	138	4	BOVCHYMOC	M14071 Bovine chym
36	23	2.0	1140	6	AX319837	AX319837 Sequence
37	23	2.0	1722	8	AT051036	U51036 Arabidopsis
38	23	2.0	1820	8	AY056403	AY056403 Arabidops
39	23	2.0	1830	8	AY063974	AY063974 Arabidops
40	23	2.0	1874	8	AY06387	AY06387 Arabidops
41	23	1.9	119	9	HUMPCPSI08	M57267 Human proch
42	22	1.9	1363	4	PIGREPA	J04601 Pig pepsino
43	22	1.9	1398	4	PIGREP	M20920 Swine pepsin
44	22	1.9	1695	8	RMASPV	X56993 R.niveus ge
45	22	1.9	1695	8	RMASPV	X56993 R.niveus ge

## ALIGNMENTS

RESULT 1  
AX088019 LOCUS 1173 bp DNA linear PAT 17-MAR-2001  
DEFINITION Sequence 1 from Patent WO0114571.  
ACCESSION AX088019  
VERSION AX088019.1 GI:13396947  
KEYWORDS

## SOURCE

ORGANISM Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovidae; Bovinae; Bos.

## REFERENCE

1 (bases 1 to 1173)  
van Rooijen, G., Keen, R.G., Boothe, J. and Shen, Y.  
Commercial production of chymosin in plants  
Patent: WO 0114571-A 1 01-MAR-2001;  
Sembiosys Genetics Inc. (CA)  
location/Qualifiers

## FEATURES

1..1173  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
1..1173  
/note="unnamed protein product"

[illegible]

OY		841	atcttgagaccggnfacgtccaaactgltggaccttagcagaacatttciaaatllag	900
Dd		841	ATCTTGAGTACCGGAGACTCCAAAGCTGGTGGAACTTAGCACGACATTCTCAACTTTAG	900
OY		901	caagcatctggagcacacacagaacacagfacggttgagtlltgcatalagattcgacaacctl	960
Dd		901	CAAGCATTTGGAGGCACACAGAACAGTAGAGGTGAAGTTTGACATFAGATTGGACAACCTT	960
OY		961	agctacatgctctacagttgtcttlttgagatcaacggccaagatgtaaccactgaaccccccc	1020
Dd		961	AGCTACATGCCCTACAGTTGTCTTTTGAAGATCAACGGCAGAAGATGTACCACCTAACCCTTC	1020
OY		1021	gccctaccagaccaagatacaagggtcttcgacacagttgattcccagaagagagaacattcc	1080
Dd		1021	GCCTTATACCAGCCAGAGATCAAGAGGTTCTTGCCACAGTAATTCAGAGTGAACACATTCC	1080
OY		1081	cagaatatgatctcttggagagatgtgtlcatlctgttgagtaactaacagcgtlcttgacaaggcc	1140
Dd		1081	CAGAAATGAGATCTTGCGAGAGTGTTCATFCTGTGAGTACTACAGCGCTTTGACAGSGCC	1140
OY		1141	aacaacctcgtttgggctagataagaacatctcga	1173
Dd		1141	AACAACCTCGTTGGGCTAGCTAATAACAATCTCGA	1173
RESULT 2				
AX088021				
LOCUS		AX088021	3957 bp	DNA
DEFINITION		Sequence 3 from Patent WO0114571.		linear
ACCESSION		AX088021		PAT 17-MAR-2001
VERSION		AX088021.1	GI:13396949	
KEYWORDS				
SOURCE		.		
ORGANISM		synthetic construct.		
REFERENCE		artificial construct.		
AUTHORS		1 (bases 1 to 3957)		
TITLE		van Rooijen, G., Keon, R.G., Boothe, J. and Shen, Y.		
JOURNML		Commercial production of chymosin in plants Patent: WO 0114571-A 3 01-MAR-2001; Sembiosys Genetics Inc. (CA) Location/Qualifiers		
FEATURES				
source		1..3957		
		/organism="synthetic construct"		
		/db_xref="taxon:32630"		
CDS		1554..2726		
		/note="Figure 2"		
		//codon_start=1		
		//transl_table=1		
		/protein_id="CAC34681.1"		
		/db_xref="gi:13396950"		
		/translation="MNFLLSPPEFAVFLCFQGYFVAVTMAHAEITRILPIYKGLRKALK EHGLLEDFLOKOQYGISKSYSGRFEVASVPLETNVLDSOYREKIYLGRPOGFVLPDFT GSSDPWPSTYCKSNACKNHORFDRPKRSSTPONTLGPPLSHYIGTSMOGLIDPTVTV SNVIDIOETVGLSTQEPEGVDVFYAEFGILLMAVPSLAASESLIPEFNMMNRHLVADD LFSEVMRNQGESMLTLGALIDPVSYYTGSLSHMVPVTVDOOMFTVDSTVISGVVACEB GCQAIDLTGSKLVGPSDDIIQOAIGAARDNOYGERPIDCDNISLYMPVVEFINGMKM YPLPSAVTSDSDQEGFCSGFENSEHSOKWILGDVFIIEYIVSVPRANNIVGLAKAI"		
BASE COUNT		1263 a	790 c	609 g 1295 t
ORIGIN				
Query Match		100.0%:	Score 1173:	DB 6; Length 3957;
Best local similarity		100.0%:	Pred. No. 0:	
Matches 1173:		Conservative 0:	Mismatches 0:	Indels 0: Gaps 0:
OY		1	atgaaactcctaagatctcttcctcttctaagcgttcccttgglttcgtaacactcgt	60
Dd		1554	ATGAAGCTCCCTTAAGCTTTCCCTTTACAGGCTTCCCTTGTTCGGCAATACTTCGT	1613
OY		61	gctgttaactaacgctgcgtgagataaacccgcatctctctctacaagaftgaagttctcgt	120
Dd		1614	GCTGTTAAGCTACGCTGCTGAGATCAACCGCATTCCTCTCTACAAAGGTAAGTCTCCGT	1673



TITLE CONJUGATED PLASMID AND MICROORGANISM CONTAINING THE SAME  
JOURNAL Patent: JP 1983032896-A 1 25-FEB-1983;  
BEPPU TERUHIKO  
COMMENT OS calf  
PN JP 1983032896-A/1  
PD 25-FEB-1983  
PI 24-AUG-1981 JP 1981131631  
PF BEPPU TERUHIKO, UOZUMI TAKESHI, NISHIMORI KATSUHIKO PC  
C07H21/04,C12N1/00,C12N15/00,(C12N1/00,C12R1:19); CC  
strandedness: Double;  
CC topology: Linear;  
CC hypothetical: No;  
CC anti-sense: No;  
CC \*source: tissue\_type=mucous membrane of forth stomach; CC  
\*source: clone=PCR 100 1;  
FH Key Location/Qualifiers  
FH CDS 1..1098  
FT Location/Qualifiers  
1..1098  
/organism="unidentified"  
/db\_xref="taxon:32644"  
BASE COUNT 252 a 327 c 300 g 219 t  
ORIGIN

Query Match 4.2%; Score 49; DB 6; Length 1098;  
Best Local Similarity 100.0%; Pred. No. 7.5e-15;  
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 656 tttctcgttaccatgacaggaatgcccagagagacatgctcacgct 704  
|||||  
Db 581 TGTTCGCTTACATGACAGGAATGGCCAGAGACATGCTCACGCT 629

RESULT 5  
E02341 1098 bp RNA linear PAT 29-SEP-1997  
LOCUS E02341  
DEFINITION cDNA sequence coding for infant bovine prorennin.  
ACCESSION E02341  
VERSION E02341.1 GI:2170576  
KEYWORDS JP 1990109984-A/1.  
SOURCE cow.  
ORGANISM Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovidae; Bovinae; Bos.  
1 (bases 1 to 1098)  
Beppu,T., Uozumi,T. and Nishimori,K.  
TITLE COMPLEX PLASMID AND MICROORGANISM CONTAINING THE SAME  
JOURNAL Patent: JP 1990109984-A 1 23-APR-1990;  
BEPPU TERUHIKO  
COMMENT OS Bovine  
PN JP 1990109984-A/1  
PD 23-APR-1990  
PI 01-DEC-1988 JP 1988302176  
PF BEPPU TERUHIKO, UOZUMI TAKESHI, NISHIMORI KATSUHIKO PC  
C12N15/59,C12N1/21;  
CC strandedness: Single;  
CC topology: Linear;  
CC hypothetical: No;  
CC anti-sense: No;  
CC \*source: tissue\_type=Stomach mucosa;  
FH Key Location/Qualifiers  
FH CDS 1..1098  
FT Location/Qualifiers  
1..1098  
/product="Infant bovine prorennin".  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
BASE COUNT 252 a 327 c 300 g 219 t

ORIGIN

Query Match 4.2%; Score 49; DB 6; Length 1098;  
Best Local Similarity 100.0%; Pred. No. 7.5e-15;  
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 656 tttctcgttaccatgacaggaatgcccagagagacatgctcacgct 704  
|||||  
Db 581 TGTTCGCTTACATGACAGGAATGGCCAGAGACATGCTCACGCT 629

RESULT 6  
AF177290 1101 bp mRNA linear MAM 24-JAN-2000  
LOCUS AF177290  
DEFINITION Bubalus arnee bubalis prothymosin mRNA, complete cds.  
ACCESSION AF177290  
VERSION AF177290.1 GI:6739579  
KEYWORDS  
SOURCE water buffalo.  
ORGANISM Bubalus bubalis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovidae; Bovinae; Bubalus.  
1 (bases 1 to 1101)  
Batish,V.K., Mukhopadhyay,U.K., Mohanty,A.K., Grover,S. and  
Kuipers,O.P.  
TITLE Direct Submission  
JOURNAL Submitted (12-AUG-1999) Molecular Biology Unit, Animal  
Biotechnology Centre, National Dairy Research Institute, G.T.Road,  
Karnal, Haryana 132001, India  
Location/Qualifiers  
1..1101  
/organism="Bubalus bubalis"  
/sub\_species="bubalis"  
/db\_xref="taxon:89462"  
1..1101  
/codon\_start=1  
/product="prothymosin"  
/protein\_id="AAF27315.1"  
/db\_xref="GI:6739580"  
CDS  
/translation="MAETTRIPLCRGSRLKALKEHGLEDFLOKQGYGVSSKSGFG  
EVASVPLNTNIDSOYRFGKIYIGTPROPTVLEFDGSSDFWVPSLYCKSNACKNHOED  
PRKSTTQNLKPLSTIRIGTSMOGIGTIDPATVSNYVDIOQTGSLSTOPEGDVFYA  
EFEDILGMAPVPSLASEYSIPEVDMNMNHLVAQVDFSYMDRNGQESMLTIGALDPSY  
YTGSIHWMPVVOQYWFVDSITISGVVACEGGCAI1D7GSKLVGSPSDLT1Q  
QAIGATONQYGEFIDCDNLSTMPYVSEIENKMYPLTPSAVTSODGFCYSGFQSEN  
RSQOMILGDVFRREYVSFDRANMLVGLAKAI"

BASE COUNT 244 a 336 c 306 g 215 t  
ORIGIN

Query Match 4.2%; Score 49; DB 4; Length 1101;  
Best Local Similarity 100.0%; Pred. No. 7.5e-15;  
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 656 tttctcgttaccatgacaggaatgcccagagagacatgctcacgct 704  
|||||  
Db 584 TGTTCGCTTACATGACAGGAATGGCCAGAGACATGCTCACGCT 632

RESULT 7  
SYNPROCAA 1115 bp DNA linear SYN 27-APR-1993  
LOCUS SYNPROCAA  
DEFINITION Synthetic bovine prothymosin A protein region.  
ACCESSION M22593 M18758  
VERSION M22593.1 GI:209139  
KEYWORDS prothymosin.  
SOURCE Synthetic DNA.  
ORGANISM Synthetic construct  
REFERENCE 1 (bases 1 to 1115)  
Wosnick,M.A., Barnett,R.W., Vicentini,A.M., Erfle,H., Elliott,R.,

TITLE Summer-Smith, M., Mantel, N. and Davies, R.W.  
Rapid construction of large synthetic genes: total chemical synthesis of two different versions of the bovine prochymosin gene

JOURNAL Gene 60, 115-127 (1987)

MEDLINE 88152494

FEATURES  
source Location/Qualifiers  
1..1115 /organism="synthetic construct"  
/db\_xref="taxon:32630"

BASE COUNT 254 a 332 c 302 g 227 t

ORIGIN 1 bp downstream of HindIII site.

Query Match 4.2%: Score 49; DB 12; Length 1115;  
Best Local Similarity 100.0%; Pred. No. 7.5e-15;  
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 656 tttctcgtttacatgacaggaatggccagagagatgctcacgct 704  
|||||  
Db 580 TGTCTCGGTTTACATGACAGATGGCCAGAGAGCATGCTCAGCT 628

RESULT 8  
104058 LOCUS 1175 bp linear PAT 02-DEC-1994  
DEFINITION Sequence 5 from Patent EP 0123928.  
ACCESSION 104058  
VERSION 104058.1 GI:591912  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 1175)  
AUTHORS Gashon, L.M., McCaman, M.T., Rice, C.W. and Sias, S.R.  
TITLE Recombinant DNA coding for a polypeptide displaying milk clotting activity  
JOURNAL Patent: EP 0123928-A2 5 07-NOV-1984;  
FEATURES  
source Location/Qualifiers  
1..1175 /organism="unknown"

BASE COUNT 263 a 356 c 318 g 238 t

ORIGIN

Query Match 4.2%: Score 49; DB 6; Length 1175;  
Best Local Similarity 100.0%; Pred. No. 7.4e-15;  
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 656 tttctcgtttacatgacaggaatggccagagagatgctcacgct 704  
|||||  
Db 649 TGTCTCGGTTTACATGACAGATGGCCAGAGAGCATGCTCAGCT 697

RESULT 9  
AR002347 LOCUS 1240 bp DNA linear PAT 04-DEC-1998  
DEFINITION Sequence 2 from patent US 5741665.  
ACCESSION AR002347  
VERSION AR002347.1 GI:3963901  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 1240)  
AUTHORS Kato, E.K. and Stuart, W. Dorsey.  
TITLE Light-regulated promoters for production of heterologous proteins in filamentous fungi  
JOURNAL Patent: US 5741665-A 2 21-APR-1998;  
FEATURES  
source Location/Qualifiers  
1..1240 /organism="unknown"

BASE COUNT 274 a 374 c 339 g 253 t

ORIGIN

Query Match 4.2%: Score 49; DB 6; Length 1240;  
Best Local Similarity 100.0%; Pred. No. 7.4e-15;  
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 656 tttctcgtttacatgacaggaatggccagagagatgctcacgct 704  
|||||  
Db 700 TGTCTCGGTTTACATGACAGATGGCCAGAGAGCATGCTCAGCT 748

RESULT 10  
BOVCHYMOA 1275 bp mRNA linear MAM 26-APR-1993  
LOCUS bovine chymosin a (rennin) mRNA.  
DEFINITION J00002  
ACCESSION J00002.1 GI:162857  
VERSION J00002.1 GI:162857  
KEYWORDS chymosin; chymosin A; rennin.  
SOURCE bovine (calf) cdna of fourth stomach mucosa mRNA.  
ORGANISM Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bovinae; Bos.  
REFERENCE 1 (bases 1 to 1275)  
AUTHORS Moir, D., Mao, J.L., Schumm, J.W., Vovis, G.F., Alford, B.L. and Taunton-Rigby, A.  
TITLE Molecular cloning and characterization of double-stranded cdna coding for bovine chymosin  
JOURNAL Gene 19, 127-138 (1982)  
MEDLINE 83054629  
COMMENT chymosin (rennin) is the major proteolytic enzyme in the fourth stomach of the unweaned calf. two chromatographically distinct forms are known and a third seems likely (see bovyhmob and bovyhmoec). this sequence has been tentatively identified as chymosin a; it differs from chymosin b by only two amino acids and from chymosin c by another amino acid. [1] argues that the different chymosins are probably polymorphic variants of a single chymosin gene.

FEATURES  
source Location/Qualifiers  
1..1275 /organism="Bos taurus"  
/db\_xref="taxon:9913"  
21..1166 /note="preprochymosin a"  
/protein\_id="AA30447.1"  
/codon\_start=1  
/db\_xref="GI:162858"  
/translation="MRCILVYLAVFALSGAETRIPIYKKSIRKALKENGLEDPLQKOYGISISKYSGFEVASVPLTNVLDISOYGKLYIGTPPOEFVLPDGTSSDWPVS IYCKSNCKNQRFDPRKSSPFGNLGRLPSLHYGTSGMOGLIGDTVYSNLIIDQRT VGLSTORPGDVFYAFERFDGILGMAVPLASVYSIPVDNMNRLVADLESYMDRN GOESMLTGLAIDPRTYTGSLHWVYTVQOYKQFTYVDSYTTISGVYVACEGGCAIILDIG TSKLVGPSSDILNTIOAIGATONQYDFEDIDCNLSYMPVVFENKMPPLTBSAVT SODGFCFTSGFQSENHSOKWILGIDVFTREYYSVDRNNMLVLAKAI"

sig\_peptide 24..68  
mat\_peptide 195..1163  
/product="chymosin a"

BASE COUNT 293 a 391 c 336 g 255 t

ORIGIN 20 bases upstream from codon 1.

Query Match 4.2%: Score 49; DB 4; Length 1275;  
Best Local Similarity 100.0%; Pred. No. 7.3e-15;  
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 656 tttctcgtttacatgacaggaatggccagagagatgctcacgct 704  
|||||  
Db 649 TGTCTCGGTTTACATGACAGATGGCCAGAGAGCATGCTCAGCT 697

RESULT 11  
E00144

LOCUS E00144 1289 bp RNA linear PAT 29-SEP-1997  
DEFINITION cDNA encoding bovine calf chymosin.  
ACCESSION E00144  
VERSION E00144.1 GI:2168443  
KEYWORDS JP 1984021392-A/1.  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 1289)  
AUTHORS Chibaarsu,E.B.  
TITLE CATTLE KIMOCIN  
JOURNAL Patent: JP 1984021392-A 1 03-FEB-1984;  
GENEX CORP  
COMMENT OS bovine calf  
PN JP 1984021392-A/1  
PD 03-FEB-1984  
PF 30-JUN-1983 JP 1983119481  
PR 01-JUL-1982 US 82 394433, 13-APR-1983 US 83 484539 PI  
CHIVARUSU EI BASURETSUTO  
PC C12N15/00,C07H21/04,C12N1/20,C12P19/34,C12P21/02,C12N15/00,  
C12R1:19);  
CC strandedness: Double;  
CC topology: linear;  
CC hypothetical: No;  
CC anti-sense: No;  
CC \*source: tissue\_type=stomach mucosa;  
FH Key Location/Qualifiers  
FT mat\_peptide 169..1113  
FT /product='bovine calf chymosin' FT CDS  
FT <1..1116 /product='precursor polypeptide'.  
FEATURES  
source Location/Qualifiers  
1..1289  
/organism='unidentified'  
/db\_xref='taxon:32644'  
BASE COUNT 311 a 388 c 329 g 260 t 1 others  
ORIGIN

Query Match 4.2%; Score 49; DB 6; Length 1289;  
Best Local Similarity 100.0%; Pred. No. 7.3e-15;  
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 656 tttctcgtttacatgacaggaatgcccagagagatgctcagct 704  
|||||  
Db 623 tttctcgtttacatgacaggaatgcccagagagatgctcagct 671

RESULT 12  
A15836 A15836 1291 bp DNA linear PAT 16-FEB-1994  
LOCUS chymosin gene.  
DEFINITION A15836  
ACCESSION A15836  
VERSION A15836.1 GI:488959  
KEYWORDS  
SOURCE .  
ORGANISM cow.  
Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovidae; Bovinae; Bos.  
REFERENCE 1 (bases 1 to 1291)  
AUTHORS Simons,A.F.M. and De Vos,W.M.  
TITLE DNA fragments, containing a lactic acid bacterium-specific  
regulator region for the expression of genes coding for normally  
heterologous proteins  
Patent: EP 0307011-A 5 15-MAR-1989;  
JOURNAL NEDERLANDS INSTITUUT VOOR ZUIVELONDERZOEK  
location/Qualifiers  
1..1291  
/organism='Bos taurus'  
/db\_xref='taxon:9913'  
BASE COUNT 323 a 382 c 328 g 257 t 1 others

ORIGIN

Query Match 4.2%; Score 49; DB 6; Length 1291;  
Best Local Similarity 100.0%; Pred. No. 7.3e-15;  
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 656 tttctcgtttacatgacaggaatgcccagagagatgctcagct 704  
|||||  
Db 609 tttctcgtttacatgacaggaatgcccagagagatgctcagct 657

RESULT 13  
BOVCHYMOM  
LOCUS BOVCHYMOM 1305 bp mRNA linear MAM 26-APR-1993  
DEFINITION bovine chymosin b (rennin) mRNA.  
ACCESSION J00003  
VERSION J00003.1 GI:162859  
KEYWORDS chymosin; chymosin B; protease; rennin.  
SOURCE bovine (calf) cDNA of fourth stomach mucosa mRNA.  
ORGANISM Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovidae; Bovinae; Bos.  
REFERENCE 1 (bases 1 to 1305)  
AUTHORS Harris,T.J.R., Lowe,P.A., Lyons,A., Thomas,P.G., Eaton,M.A.W.,  
Mallican,T.A., Patel,T.P., Bose,C.C., Carey,N.H. and Doel,M.T.  
TITLE molecular cloning and nucleotide sequence of cDNA coding for calf  
preprochymosin  
JOURNAL Nucleic Acids Res. 10, 2177-2187 (1982)  
MEDLINE 82221400  
COMMENT chymosin is the major proteolytic enzyme in the fourth stomach of  
the unweaned calf. Two chromatographically different forms, a and  
b, of the enzyme and its precursor are known and a third form seems  
likely (see bovychnoma, bovychnoc). This sequence has tentatively  
been identified as coding for preprochymosin b. Sequence comparison  
indicates that the precursors for chymosins a and b differ by only  
two amino acids, and for b and c by only four amino acids.

FEATURES  
source Location/Qualifiers  
1..1305  
/organism='Bos taurus'  
/db\_xref='taxon:9913'  
<1..1305  
/product='chymob mRNA'  
26..1171  
/note='preprochymosin b'  
/codon.start=1  
/protein.id='AAA30448.1'  
/db\_xref='GI:162860'  
/translation='MRLVYLLAVFALSQAEITRIPYKSKSRALKRHLDEL  
QKQVYGISKRSYSGFEVAVPLNTYLDSTQYFGKLYLGTPOEFTVLFDTSSDPWVS  
IYCKSNACKNHQRFDPKRSSTFQMLGRPLIHVGTGSMQILGDTVTSNIVDIQT  
VGLSTQEPGDVFTAEFDGILGMAYPSLAESYSPVDSMMNRLLVADDFSVYMDN  
GOESMLTGLAINPSYTGSLHWWVVTVOQWOFVDSVTTSIGVVAACEGCAALIDTG  
TSKRGSSPILNITIOALIGATONQGFEDIDCNLSYMPVVEPIENKMPPLPSAVT  
SDDGFGCTSGFQSEHSQKMTLGDVFTIREYTSYFRANNLVGLAKAI'  
29..73  
/note='presequence'  
mat\_peptide 200..1168  
/product='chymosin b'  
BASE COUNT 305 a 393 c 340 g 267 t  
ORIGIN pst-1 site.

QY 656 tttctcgtttacatgacaggaatgcccagagagatgctcagct 704  
|||||  
Db 654 tttctcgtttacatgacaggaatgcccagagagatgctcagct 702



RESULT 14  
E00042 1460 bp RNA linear PAT 29-SEP-1997  
LOCUS E00042  
DEFINITION DNA coding of pre-pro rennin.  
ACCESSION E00042  
VERSION E00042.1 GI:2168348  
KEYWORDS JP 1982141287-A/1.  
SOURCE unidentified.  
ORGANISM unclassified.  
REFERENCE 1 (bases 1 to 1460)  
AUTHORS Baanadetsute,R.A., Jient,M., Donarudo,T.M., Arison,T.R. and Jierarudo,F.B.  
TITLE RENININ, PREPRORENININ OR PRORENININ GENE OBTAINED FROM RECOMBINED DNA MATERIAL AND LIVE CELL CONTAINING GENE  
JOURNAL Patent: JP 1982141287-A 1 01-SEP-1982;  
COMMENT KORABORATEIBU RES INC  
OS call  
PN JP 1982141287-A/1  
PD 01-SEP-1982  
PF 14-JAN-1982 JP 1982003556  
PR 16-JAN-1981 US 81 225717, 01-DEC-1981 US 81 325481 PI  
BAANADETSUTE RABUTSUTSUKII ARUFUODO, JIENT MAO, PI DONARUDO TEIRAA MOJIA,  
PI ARISON TAUNTON RIGUBII, JIERARUDO FURANSHISU BUBOISU PC  
C12N1/00,C07G7/00,C07H21/04,C12N15/00,C12P21/00//C12N1/18, PC  
C12N1/20,  
PC C12P19/34,C12R1/19,C12R1/865;  
CC strandedness: Double;  
CC topology: Linear;  
CC hypothetical: No;  
CC anti-sense: No;  
CC \*source: tissue\_type=stomach;  
FH key Location/Qualifiers  
FT CDS 205..1350  
FT /product='pre-pro rennin'  
FT mat\_peptide 205..1347  
FT /product='pre-pro rennin'  
FT CDS 253..1350  
FT /product='pro rennin'  
FT mat\_peptide 253..1347  
FT /product='pro rennin'  
FT CDS 278..1350  
FT /product='pro rennin'  
FT mat\_peptide 278..1347  
FT /product='rennin'  
FEATURES  
source Location/Qualifiers  
1..1460  
/organism='unidentified'  
/db\_xref='taxon:32644'  
BASE COUNT 328 a 440 c 398 g 294 t  
ORIGIN  
Query Match 4.2%; Score 49; DB 6; Length 1460;  
Best Local Similarity 100.0%; Pred. No. 7.2e-15;  
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 656 tggttcggtttacatgacaggaatgccaagagagcatgctcacgct 704  
|||||  
Db 833 tcttctcggtttacatgacaggaatgccaagagagcatgctcacgct 881  
RESULT 15  
E00295 1460 bp RNA linear PAT 29-SEP-1997  
LOCUS E00295  
DEFINITION cDNA encoding pre-prorennin A.  
ACCESSION E00295  
VERSION E00295.1 GI:2168583  
KEYWORDS JP 1985058077-A/3.  
SOURCE cow.  
ORGANISM Bos taurus

REFERENCE 1 (bases 1 to 1460)  
AUTHORS Baanadetsute,R.A., Jient,M., Donarudo,T.M., Arison,T.R., Robaato,J.N., Jien,I.M., Donarudo,T.M. and Kurisutofuna,G.G.  
TITLE USE OF GAL YEAST PROMOTOR  
JOURNAL Patent: JP 1985058077-A 3 04-APR-1985;  
COMMENT KORABORATEIBU RES INC  
OS Bovine  
PN JP 1985058077-A/3  
PD 04-APR-1985  
PF 28-FEB-1984 JP 1984035472  
PR 28-FEB-1983 US 83 470911  
PI DEBUTSUDO BOTSTOSUTPEIN, RONARUDO UETN DEIBUISU, PI JIERARUDO RARUFU FJINKU,  
PI ARISON TAUNTON RIGUBII, ROBAATO JIENTORII NOURUTON, JIENT I MAO, PI DONARUDO TEIRAA MOA, KURISUTOFUNA GONSUDOFURII GOFU PC  
C12N15/00,C07H21/04,C07K13/00,C12N1/16,C12P21/02,C12N1/16, PC  
C12R1/865),  
PC (C12P21/02,C12R1/865);  
CC strandedness: Double;  
CC topology: Linear;  
CC hypothetical: No;  
CC anti-sense: No;  
CC \*source: tissue\_type=stomach mucosa;  
FH key Location/Qualifiers  
FT CDS 205..1350  
FT /product='pre-prorennin A'  
FT mat\_peptide 205..1347  
FT /product='pre-prorennin A'  
FT mat\_peptide 253..1347  
FT /product='Prorennin'.  
FEATURES  
source Location/Qualifiers  
1..1460  
/organism='Bos taurus'  
/db\_xref='taxon:9913'  
BASE COUNT 327 a 437 c 400 g 296 t  
ORIGIN  
Query Match 4.2%; Score 49; DB 6; Length 1460;  
Best Local Similarity 100.0%; Pred. No. 7.2e-15;  
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 656 tggttcggtttacatgacaggaatgccaagagagcatgctcacgct 704  
|||||  
Db 833 tcttctcggtttacatgacaggaatgccaagagagcatgctcacgct 881

Search completed: August 1, 2002, 01:41:47  
Job time: 16797 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 21:15:05 ; Search time 791.3 Seconds  
(without alignments)  
2545.104 Million cell updates/sec

Title: US-09-643-755B-1

Perfect score: 1173  
Sequence: 1 atgaacttccttaagctctt.....gctagctaaagcaatctga 1173

Scoring table:

OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 1736436 seqs, 858457221 residues

Word size: 0

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database: N\_Geneseq\_032802.\*

```
1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	1173	100.0	22 AAS00569	Bovine pre-pro-chy
2	1173	100.0	22 AAS00570	Bovine phaseolin p
3	49	4.2	1098 11 AAN30063	Sequence encoding
4	49	4.2	1098 11 AAO04683	Sequence encoding
5	49	4.2	1175 5 AAN40295	Prochymosin
6	49	4.2	1175 13 AAO20949	Prochymosin (prore
7	49	4.2	1210 12 AAQ14051	Renin gene. Synt
8	49	4.2	1240 16 AAT03006	Chymosin open read
9	49	4.2	5 AAN40055	Sequence of prochy

10	49	4.2	1289 4 AAN30022	Sequence of veal c
11	49	4.2	1291 10 AAN91157	Cloned sequence of
12	49	4.2	1314 4 AAN30049	CDNA sequence corr
13	49	4.2	1460 3 AAN20043	Pre-prorenin-A ge
14	49	4.2	1460 5 AAN40180	Sequence of recomb
15	49	4.2	2727 10 AAN91188	BamHI/SalI insert
16	49	4.2	2733 20 AAO206463	2.7 Kbp HindIII fr
17	49	4.2	2982 10 AAN91185	BamHI insert from
18	38	3.2	1290 4 AAN30209	Sequence of prepro
19	35	3.0	637 14 AAO49459	Prochymosin gene 5
20	34	2.9	1096 20 AAV83966	DNA encoding the f
21	33	2.8	179 5 AAN40296	Sequence coding fo
22	33	2.8	186 13 AAO20950	DNA coding for a r
23	33	2.8	819 20 AAV83967	DNA encoding His-c
24	32	2.7	107 12 AAO14777	Renin - casein co
25	29	2.5	1460 3 AAN20043	Pre-prorenin-A ge
26	29	2.5	1460 5 AAN40180	Sequence of recomb
27	23	2.0	1083 21 AAC44658	Arabidopsis thalia
28	23	2.0	1140 24 AAS97159	Human aspartyl pro
29	23	2.0	1814 21 AAC37875	Arabidopsis thalia
30	20	1.7	828 11 AAO03224	DNA fragment of pa
31	20	1.7	846 11 AAO03223	DNA fragment of pa
32	20	1.7	1104 9 AAN80001	Optimised prochymo
33	20	1.7	1143 5 AAN40214	Gene encoding (pre
34	20	1.7	11624 19 AAV52850	Human eyal gene co
35	19	1.6	231 23 AAS57374	cDNA #50 encoding
36	19	1.6	689 19 AAV59757	Human secreted pro
37	19	1.6	1353 21 AAZ39862	Plasmodium ovale p
38	19	1.6	1751 19 AAV59610	Human secreted pro
39	19	1.6	1774 20 AAZ42126	Human normal blad
40	19	1.6	3362 18 AAV06278	Human putromycin-se
41	19	1.6	3747 18 AAV06250	Human putromycin-se
42	19	1.6	4049 22 AAH02878	Human shear stress
43	19	1.6	4170 22 AAD09546	Human protease pro
44	18	1.5	158 21 AAC21764	Human secreted pro
45	18	1.5	430 22 AAK60501	Human immune/haema

#### ALIGNMENTS

RESULT 1	
AAS00569	standard; DNA: 1173 Bp.
ID	AAS00569 standard; DNA: 1173 Bp.
AC	AAS00569;
XX	
DT	14-MAY-2001 (first entry)
DE	Bovine pre-pro-chymosin DNA sequence.
XX	
KW	Chymosin; transcription regulator; terminator sequence; soybean; corn;
KW	pre-pro-chymosin; rape seed; sunflower; cotton; tobacco; alfalfa; wheat;
KW	barley; oats; sorghum; Arabidopsis thaliana; potato; flax; linseed; rice;
KW	safflower; oil palm; ground nut; Brazil nut; coconut; castor; coriander;
KW	squash; jojoba; ds.
OS	Bos sp.
XX	
FH	Key
FT	1..1173
FT	CDS
FT	
FT	
FT	sig_peptide
FT	1..78
FT	/tag= b
FT	/product= "Bovine chymosin"
FT	
FT	misc_feature
FT	79..201
FT	/tag= c
FT	/note= "Pro sequence"
FT	202..1170
FT	/tag= d
FT	/product= "Mature bovine chymosin"
FT	
XX	
PN	MO200114571-A1.

XX 01-MAR-2001.  
PD  
XX  
PF 23-AUG-2000; 2000WO-CA00975.  
XX  
PR 23-AUG-1999; 99US-0378696.  
XX  
PA (SEMB-) SEMBIOSYS GENETICS INC.  
XX  
PI Van Rooijen G, Keon RG, Boothe J, Shen Y;  
XX  
DR WPI; 2001-226621/23.  
DR P-PSDB; AAU00536.  
XX  
PT Producing chymosin in seeds of plants such as rice, flax, rape seed, by  
PT transforming plant cell with a nucleic acid encoding chymosin operably  
PT linked to transcription regulator and terminator sequences -  
XX  
XX  
XX Claim 9; Fig 1; 56pp; English.  
XX

The sequence represents a DNA which encodes a bovine chymosin polypeptide. Chymosin can be produced in a plant seed through introduction of a chimeric nucleic acid molecule, comprising a nucleic acid sequence encoding a chymosin polypeptide operatively linked to a transcription regulator and terminator sequences, into a plant cell. The sequences are useful for producing plant seeds, in particular seeds of soybean, rape seed, sunflower, cotton, corn, tobacco, alfalfa, wheat, barley, oats, sorghum, *Arabidopsis thaliana*, potato, flax/linseed, safflower, oil palm, groundnut, Brazil nut, coconut, castor, coriander, squash, jojoba and rice.

Sequence 1173 BP; 299 A; 308 C; 262 G; 304 T; 0 other;

Query Match	100.0%	Score 1173;	DB 22;	Length 1173;
Best Local Similarity	100.0%	Pred. No. 0;		
Matches 1173;	Conservative	0;	Mismatches	0;
			Indels	Gaps
QY 1 atgaacttccttaagcttcctccctctctacagcttccttccttggttcgtaataacttcgtt				
Db 1 atgaacttccttaagcttcctccctctctacagcttccttccttggttcgtaataacttcgtt				
QY 61 gctgttaccacagctctgtatagatacaaccgcatctcctctctaaagaatgactctccgt				
Db 61 gctgttaccacagctctgtatagatacaaccgcatctcctctctaaagaatgactctccgt				
QY 121 aagcgctgaaaggaacatgagactcttagaagactctcttcgacagaaacaaagatggtcacc				
Db 121 aagcgctgaaaggaacatgagactcttagaagactctcttcgacagaaacaaagatggtcacc				
QY 181 agcagaagaatctccggtctgtgtagaagtgtctagagtgccacttaaccataactcttgat				
Db 181 agcagaagaatctccggtctgtgtagaagtgtctagagtgccacttaaccataactcttgat				
QY 241 agtcaatactttvggagaagatactcaaccgcggaaccgcgcctcaagatctacogtctcttt				
Db 241 agtcaatactttvggagaagatactcaaccgcggaaccgcgcctcaagatctacogtctcttt				
QY 301 gatacttgttccctctgactctctggtgttcctctatacttaactgcaagagcaatgcctgcaag				
Db 301 gatacttgttccctctgactctctggtgttcctctatacttaactgcaagagcaatgcctgcaag				
QY 361 aaccaccaagaatctcgatcccgagaagaagtgtgcacacttccagaacttaaggcaaaccttg				
Db 361 aaccaccaagaatctcgatcccgagaagaagtgtgcacacttccagaacttaaggcaaaccttg				
QY 421 tctatactactacgtgtacagtagacatgcagaagaagatcttaagcctatataccgctacgttc				
Db 421 tctatactactacgtgtacagtagacatgcagaagaagatcttaagcctatataccgctacgttc				
QY 481 tccaacatttgtgacattcaacagacagttagaacttagaaccacaagaacaggtgatgac				
Db 481 tccaacatttgtgacattcaacagacagttagaacttagaaccacaagaacaggtgatgac				

QY	541	ttcaactatgcgaatctcgatggaatcccttggtagtgcatacccatcgctcgcgttcagag	600
Db	541	ttcaactatgcgaatctcgatggaatcccttggtagtgcatacccatcgctcgcgttcagag	600
QY	601	tactcgataccgctgtttggacaacatgtatgaacccgacacactgtatgctcaaatgtttc	660
Db	601	tactcgataccgctgtttggacaacatgtatgaacccgacacactgtatgctcaaatgtttc	660
QY	661	tcggtttacatggaaggaatgycagaggaagacatgctacgcttggagctattgatcca	720
Db	661	tcggtttacatggaaggaatgycagaggaagacatgctacgcttggagctattgatcca	720
QY	721	tcctactaacacggatctcttcttaccggtgttcaagtcacgtgtcagcagatcttgcgaatc	780
Db	721	tcctactaacacggatctcttcttaccggtgttcaagtcacgtgtcagcagatcttgcgaatc	780
QY	781	actgtggacagtgtaaccaatcagcgggtgtgtctgtgtgtgtgtgtgtgtgtgtgtgtgt	840
Db	781	actgtggacagtgtaaccaatcagcgggtgtgtctgtgtgtgtgtgtgtgtgtgtgtgtgt	840
QY	841	atcttggataccggtatcgtccaaagctgtgtcggacctagcagcagacattctcaacattcag	900
Db	841	atcttggataccggtatcgtccaaagctgtgtcggacctagcagcagacattctcaacattcag	900
QY	901	caagcatttgaggccaacagaacccagtagtcggtgagtttgacatagatttgcgacaacct	960
Db	901	caagcatttgaggccaacagaacccagtagtcggtgagtttgacatagatttgcgacaacct	960
QY	961	agctacatgctccacagttgtcttcttggatcaacggacaagaatgtatccactgacccctcc	1020
Db	961	agctacatgctccacagttgtcttcttggatcaacggacaagaatgtatccactgacccctcc	1020
QY	1021	gcctataccagccaggaatcaagggtctctgcacagtcggatttccagggttagaaccattcc	1080
Db	1021	gcctataccagccaggaatcaagggtctctgcacagtcggatttccagggttagaaccattcc	1080
QY	1081	cagaatgtgatacttggagatgtgttcatcttcgtgagtaatacagcgtctcttgcagggcc	1140
Db	1081	cagaatgtgatacttggagatgtgttcatcttcgtgagtaatacagcgtctcttgcagggcc	1140
QY	1141	aacaacctcgttgggctagctaaagaacattcga	1173
Db	1141	aacaacctcgttgggctagctaaagaacattcga	1173
RESULT 2			
AAS00570			
ID	AAS00570 standard; DNA; 3957 BP.		
XX			
AC	AAS00570;		
XX			
DT	14-MAY-2001 (first entry)		
XX			
DE	Bovine phaseolin promoter pre-pro-chymosin-phaseolin terminator.		
XX			
KW	Chymosin; transcription regulator; terminator sequence; soybean; corn;		
KW	pre-pro-chymosin; rape seed; sunflower; cotton; tobacco; alfalfa; wheat;		
KW	barley; oats; sorghum; Arabidopsis thaliana; potato; flax; linseed; rice;		
KW	saflower; oil palm; ground nut; Brazil nut; coconut; castor; coriander;		
KW	squash; jojoba; ds; phaseolin; promoter; terminator; mutant; French bean.		
XX			
OS	Chimeric - Bos sp.		
XX			
XX	Chimeric - Phaseolus vulgaris.		
XX			
FH	Key		
FT	promoter		
FT	Location/Qualifiers		
FT	1..1553		
FT	/tag= a		
FT	/note= "Phaseolin promoter"		
FT	1554..2726		
FT	/tag= b		
FT	/product= "Bovine pre-pro-chymosin"		
FT	2727..3957		
FT	terminator		

```
FT      /*tag= C
FT      /note= "Phaseolin terminator"
XX      MO200114571-A1.
XX      01-MAR-2001.
XX      23-AUG-2000; 2000MO-CA00975.
XX      23-AUG-1999; 99US-0378696.
XX      (SEMB-) SEMBIOSYS GENETICS INC.
XX      Van Rooijen G, Keon RG, Boothe J, Shen Y;
XX      WPI: 2001-226621/23.
XX      P-PSDB: AAU00536.
XX      Producing chymosin in seeds of plants such as rice, flax, rape seed, by
PT      transforming plant cell with a nucleic acid encoding chymosin operably
PT      linked to transcription regulator and terminator sequences -
XX
PS      Example 1; Fig 2; 56pp; English.
XX
CC      The sequence represents a chimeric polynucleotide comprising a
CC      pre-pro-chymosin, a phaseolin promoter and a phaseolin terminator.
CC      Chymosin can be produced in a plant seed through introduction of a
CC      chimeric nucleic acid molecule, comprising a nucleic acid sequence
CC      encoding a chymosin polypeptide operatively linked to transcription
CC      regulator and terminator sequences, into a plant cell. The sequences are
CC      useful for producing plant seeds, in particular seeds of soybean, rape
CC      seed, sunflower, cotton, corn, tobacco, alfalfa, wheat, barley, oats,
CC      sorghum, Arabidopsis thaliana, potato, flax/linseed, safflower, oil palm,
CC      groundnut, Brazil nut, coconut, castor, coriander, squash, jojoba and
CC      rice.
XX
SQ      Sequence 3957 BP; 1263 A; 790 C; 609 G; 1295 T; 0 other;

Query Match      100.0%; Score 1173; DB 22; Length 3957;
Best Local Similarity 100.0%; Pred. NO. 0;
Matches 1173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 atgaattccttaagctcttcctccttctaagcgttccttcgttgcgtcaatactcgtt 60
DB      1554 atgaattccttaagctcttcctccttctaagcgttccttcgttgcgtcaatactcgtt 1613
QY      61 gctgttaactacgctgctgagatcacccgcattcctctctacaagaagtaagtcctcgt 120
DB      1614 gctgttaactacgctgctgagatcacccgcattcctctctacaagaagtaagtcctcgt 1673
QY      121 aagggcgtgaaggaaactggaactttagaagaactctcttgcaaaaacacagatgcatc 180
DB      1674 aagggcgtgaaggaaactggaactttagaagaactctcttgcaaaaacacagatgcatc 1733
QY      181 agcagaactgaactccggttcgttgaagtgttgtagcgtgcacccttaaccaactctgat 240
DB      1724 agcagaactgaactccggttcgttgaagtgttgtagcgtgcacccttaaccaactctgat 1793
QY      241 agtcaatactttgggaagatctaactcggaaaccgcgcctcaagaagttcacgttcttt 300
DB      1794 agtcaatactttgggaagatctaactcggaaaccgcgcctcaagaagttcacgttcttt 1853
QY      301 gatactggttactcgtgacttggttggcttactatcactactgaagaagaatgctgcgaag 360
DB      1854 gatactggttactcgtgacttggttggcttactatcactactgaagaagaatgctgcgaag 1913
QY      361 aaccaccaaagatcgcagcagaaagtcgtccacacttccagaacttgaagcacaaccttg 420
DB      1914 aaccaccaaagatcgcagcagaaagtcgtccacacttccagaacttgaagcacaaccttg 1973
QY      421 tctataactacggtacaggttagcatgcaaggaatccttaagctatgatacgttcactgtc 480
DB      tctataactacggtacaggttagcatgcaaggaatccttaagctatgatacgttcactgtc 480
```

```
DB      1974 tctataactacggtacaggttagcatgcaaggaatccttaagctatgatacgttcactgtc 2033
QY      481 tccaattgttggaactttaaacaagatgagacttagacacccaagaacaggtgctgc 540
DB      2034 tccaattgttggaactttaaacaagatgagacttagacacccaagaacaggtgctgc 2093
QY      541 ttaacctatgacaagaatcgcagtccttggtatgtgcaataccatgcgtcgtcagaag 600
DB      2094 ttaacctatgacaagaatcgcagtccttggtatgtgcaataccatgcgtcgtcagaag 2153
QY      601 taactgataacctgtgttgaacaacatgataagccgaaccactagatgctcaagactgttc 660
DB      2154 taactgataacctgtgttgaacaacatgataagccgaaccactagatgctcaagactgttc 2213
QY      661 tcggtttacatggaacgaatggtccagagagcatgcttaagcttggaaactattatcca 720
DB      2214 tcggtttacatggaacgaatggtccagagagcatgcttaagcttggaaactattatcca 2273
QY      721 tctactacaaggaatccttcaactgtgttcagatcactgttcagcagatgctgcgaatc 780
DB      2274 tctactacaaggaatccttcaactgtgttcagatcactgttcagcagatgctgcgaatc 2333
QY      781 actgttgacagtgctcaacatcagcgtgtgtgtgtgtgcatgtgaaagtgtatgctcaagct 840
DB      2334 actgttgacagtgctcaacatcagcgtgtgtgtgtgtgcatgtgaaagtgtatgctcaagct 2393
QY      841 atcttggataccggtgacgtccaaagctgtgtcgaacctagcagaagcatcttcaaatcag 900
DB      2394 atcttggataccggtgacgtccaaagctgtgtcgaacctagcagaagcatcttcaaatcag 2453
QY      901 caagcatttgaggccacaagaaccagatcaggtgagttgacatagatgcgacaacct 960
DB      2454 caagcatttgaggccacaagaaccagatcaggtgagttgacatagatgcgacaacct 2513
QY      961 agctatactgctacagtgcttctttagatcaacggaagatgtaccactgaccccttc 1020
DB      2514 agctatactgctacagtgcttctttagatcaacggaagatgtaccactgaccccttc 2573
QY      1021 gctctacagcagaagatgaagggttctgcacagtggaattcagaagttagaacaatttc 1080
DB      2574 gctctacagcagaagatgaagggttctgcacagtggaattcagaagttagaacaatttc 2633
QY      1081 cagaatgagatcttggaagatgtgttcattcgttgagtaactacagcgtcttgcagggcc 1140
DB      2634 cagaatgagatcttggaagatgtgttcattcgttgagtaactacagcgtcttgcagggcc 2693
QY      1141 aacaacctcgttgggtagctaaagaatctga 1173
DB      2694 aacaacctcgttgggtagctaaagaatctga 2726

RESULT      3
AAN30063
ID      AAN30063 standard; cDNA; 1098 BP.
XX
AC      AAN30063;
XX
DE      14-JUN-1992 (first entry)
XX
KW      Sequence of prorennin cDNA in PCR 10001.
XX
OS      Bos taurus.
XX
FH      Key
FT      CDS
XX      Location/Qualifiers
XX      EP73029-A.
XX      1..1098
XX      /*tag= a
XX      02-MAR-1983.
XX
```









SO Sequence 1289 BP; 310 A; 391 C; 327 G; 260 T; 1 other;

## Query Match

Best Local Similarity 4.2%; Score 49; DB 4; Length 1289;  
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 656 tttctcgtttacatgacaggaatggccagagagatgtctacgct 704  
|||||  
Db 623 tttctcgtttacatgacaggaatggccagagagatgtctacgct 671

## RESULT 11

AAN91157  
ID AAN91157 standard; DNA; 1291 BP.

XX AAN91157;

DT 07-JUN-1990 (first entry)

XX Cloned sequence of (pro)chymosin.

DE Lactic acid bacteria; cheese; Streptococcus cremoris SK112; chymosin;  
KM proteinase; PSK112; ss.

XX Key location/Qualifiers  
FH precursor\_RNA 29..1123  
FT /\*tag= a  
FT /product=;prochymosin

XX NL8701378-A.

XX 02-JAN-1989.

XX 12-JUN-1987; 87NL-0001378.

XX 12-JUN-1987; 87NL-0001378.

XX (NEZU-) NEDERL INS ZUIVELON.

XX Simons AFM, De Vos WM.

XX WPI; 1989-030097/04.

XX P-PSDB; AAP94144.

XX DNA fragment having region specific for lactic acid bacteria -  
PT is contained in plasmid in microorganism used in prodn. of  
protein and food prodn. eg cheese.

XX Disclosure; fig 2; 43pp; Dutch.

XX The DNA encoding prochymosin can be cloned into a plasmid (esp. from  
CC S. cremoris SK112) and used to produce large amts of the protein by  
CC recombinant DNA techniques. This could overcome the shortage of  
CC prochymosin due to shortage of calf stomachs and increasing cheese  
CC prodn. Prochymosin is also used in prodn. of yoghurt, butter and  
CC buttermilk. See also AAN91158-N91160.

XX Sequence 1291 BP; 323 A; 385 C; 326 G; 257 T; 0 other;

## Query Match

Best Local Similarity 4.2%; Score 49; DB 10; Length 1291;  
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 656 tttctcgtttacatgacaggaatggccagagagatgtctacgct 704  
|||||  
Db 609 tttctcgtttacatgacaggaatggccagagagatgtctacgct 657

## RESULT 12

AAN30049  
ID AAN30049 standard; cDNA; 1314 BP.

XX AAN30049;

DT 14-JUN-1992 (first entry)

DE cDNA sequence corresponding to one of the allelic forms (B) of  
DE bovine preprochymosin.

XX Chymosin; enzyme; rennet; cheese; ss.

XX Bos taurus.

XX Key location/Qualifiers  
FH sig\_peptide 24..40  
FT /\*tag= a  
FT mat\_peptide 41..1169  
FT /\*tag= b

XX EP77109-A.

XX 20-APR-1983.

XX 13-OCT-1982; 82EP-0201272.

XX 14-OCT-1981; 81GB-0031004.

XX (UNIL ) UNILEVER NV.

XX Maat J, Verrips CT, Ledebuer AM, Edens L;

XX WPI; 1983-39656K/17.

XX P-PSDB; AAP30086.

XX DNA molecules comprising genes for preprochymosin - used to  
PT transform microorganisms to give strain producing the  
PT prepro-enzyme and its allelic and maturation forms

XX Claim 2; Fig 1; 53pp; English.

XX Preprochymosin is an intermediate (via prochymosin and  
CC pseudochymosin) for the enzyme chymosin, which is the essential milk-  
CC clotting component of rennet and is used in cheese manufacture.

CC AAN30049 corresp. to mRNA isolated from the fourth stomach of a  
CC preunlulant calf (abomasum, Frisian cow).

XX Sequence 1314 BP; 309 A; 398 C; 338 G; 269 T; 0 other;

## Query Match

Best Local Similarity 4.2%; Score 49; DB 4; Length 1314;  
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 656 tttctcgtttacatgacaggaatggccagagagatgtctacgct 704  
|||||  
Db 652 tttctcgtttacatgacaggaatggccagagagatgtctacgct 700

## RESULT 13

AAN20043  
ID AAN20043 standard; DNA; 1460 BP.

XX AAN20043;

DT 16-DEC-1992 (first entry)

XX Pre-prorennin-A gene DNA sequence.

XX Pre-pro-remnin; rennin; prorennin; enzyme; EC-3.4.23.4; chymosin;  
KM protease; milk-clotting enzyme; ss.

XX Bos taurus.

XX Key location/Qualifiers

FT	CDS		205..1350	**tag=	a
FT			/label-		
FT			/note= "pre-prorennin-A gene"		
XX					
PX	GB2091271-A.				
PD	28-JUL-1982.				
PE	15-JAN-1982;	82GB-0001120.			
PX					
PR	01-DEC-1981;	81US-0325481.			
PR	16-JAN-1981;	81US-0225717.			
PA	(COLB ) COLLABORATIVE RES INC.				
PX					
PI	Alford BL, Mao J, Moir DT;				
DR	WPI; 1982-62028E/30 (62028E).				
DR	P-PSDB; AAP20038.				
PX					
PT	Transformed cells producing rennin and its precursors - contg.				
PT	appropriate recombinant DNA material				
PX					
PS	Disclosure; Table I; 39pp; English.				
PX					
CC	Bases 1-204 and 1351 to 1460 are attached to the pre-prorennin but				
CC	can be removed and are not essential to use of the gene in				
CC	expression. The gene may be ligated into plasmid pCGE21 and				
CC	expressed in E. coli. The resulting expressed enzyme is a well				
CC	known milk-clotting enzyme used in cheese-making.				
SQ	Sequence 1460 BP; 329 A; 440 C; 397 G; 294 T; 0 other:				
OY	Query Match	4.2%; Score 49; DB 3; Length 1460;			
	Best Local Similarity	100.0%; Pred.No. 1e-14;			
	Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
Dy	656 tgcctcggttacatgaccaggaattgccaggagaagcatgtcacagt 704       833 tgcttcggttacatgacagcaaatgtcccgagagagcatgtccaacct 881				
ID	AAN40180 standard; cDNA; 1460 BP.				
AC	AAN40180;				
DT	25-JAN-1992 (first entry)				
DE	Sequence of recombinant CGF4 carrying the rennin coding sequence.				
KW	Yeast expression vector; GAL1 promoter; Saccharomyces cerevisiae; ss.				
XK					
OS	Bos taurus.				
PX					
FH	Key Location/Qualifiers				
FT	CDS	205..1350			
FT		/tag=a			
PX					
PN	GB2137208-A.				
PX					
PD	03-OCT-1984.				
PX					
PF	28-FEB-1984;	84GB-0405129.			
PX					
PR	28-FEB-1983;	83US-0470911.			
PX					
PA	(COLB ) COLLABORATIVE RES INC.				
PX					

Pt	Borststein D, Davis RM, Flink GR, Taunton-Rigby A,		Knowledge RG;
Pf	Mao JI, Moir DT, Goff CG;		
Xx			
Dk	WPI; 1984-245517/40.		
Dr	P-PsDB; AAP40218.		
Xx			
Pt	DNA segment contg. GAL1 promoter linked to gene - useful for direction of expression of the gene in yeast cell		
Xx			
Ps	Example: Table 4, Page 21-23; 35pp; English.		
Xx			
Cc	The inventors claim a DNA segment contg. GAL1 promoter linked to gene - useful for direction of expression of the gene in yeast cell.		
Cc	The recombinant material carrying a GAL1 promoter of the yeast galactokinase gene may be used in expressing a desired protein, esp. bovine growth hormone, interferon, prolamin or preprolamin, in the yeast cell. Strains of Saccharomyces cerevisiae producing the polypeptides are produced. Yeast strains deposited as ATCC 20663, CC 20661, 20662 and 20663, strain designations Cgy 196, 457, 461 and 528, resp. are new.		
Cc			
Sq	Sequence 1460 BP; 328 A; 440 C; 399 G; 293 T; 0 other:		
Gy	tgtctcgattacatgcagcaaggaaagtggccagaagacgatgtcacgct 704       tttgttcggttacatgcagcaaggaaagtggccagaagacgatgtcacgct 881	4.2%; Score 49; DB 5; Length 1460; Best Local Similarity 100.0%; Pred. No. 1e-14; Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0.	
Db	833 ttttgctggttacatgcagcaaggaaagtggccagaagacgatgtcacgct 881		
An	RESULT 15		
Aa	AAN91188		
ID	AAN91188 standard; DNA; 2727 BP.		
Ac	AAN91188;		
Df	15-JUN-1990 (first entry)		
Xx			
De	BamHI/SalI insert of Kluyveromyces plasmid PAB309.		
Kw	Kluyveromyces; pBM100PC; chymosin; tissue plasminogen activator.		
Oz	Kluyveromyces lactis.		
Fh	Key Location/Qualifiers		
Ft	CDS 409..1781		
Ft	/tag= a		
Xx			
Pn	EP301670-A.		
Xx			
Pd	01-JAN-1989.		
Pf	28-JUL-1988; 88BP-0201632.		
Pr	28-JUL-1987; 87US-0078539.		
Pa	(KONN ) GIST-BROCADES NV.		
Pi	van den Berg JA, van Ooyen AJJ, Rietveld K;		
Dr	WPI; 1989-033565/05.		
DR	P-PsDB; AAP94376.		
Pt	Kluyveromyces host cells for producing polypeptide(s) - used for highly efficient produ. of eg chymosin tissue plasminogen activator or human serum albumin.		
Xx	Disclosure; : 56pp; English.		
PS			
Cc	BamBI/Sali insert, incorporated into plasmids PAB309 with		

CC promoters, terminators and a G418 resistance marker fused to a ADHI  
CC promoter from *S.cerevisiae*.  
CC The Kluyveromyces expression systems provide highly efficient secretion  
CC and processing of a wide variety of proteins.  
CC Sequences identical to those published in EP301669.  
XX  
SQ Sequence 2727 BP; 752 A; 654 C; 572 G; 749 T; 0 other;

Query Match 4.2%; Score 49; DB 10; Length 2727;  
Best Local Similarity 100.0%; Pred. No. 9.8e-15;  
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 656 tgttctcggttacatgacaggaatgcccagagagacatgctcacgct 704  
|||||  
Db 1268 tgttctcggttacatgacaggaatgcccagagagacatgctcacgct 1316

Search completed: August 1, 2002, 01:57:49  
Job time: 16964 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 18:19:35 ; Search time 5849.81 seconds  
(without alignments)  
2706.400 Million cell updates/sec

Title: US-09-643-755b-1

Perfect score: 1173

Sequence: 1 atgaacttccttaagctctt.....ggctagctaagcaatcgtg 1173

Scoring table: OLIGO\_NUC

Gapop 60.0 , Gapext 60.0

Searched: 13736207 seqs, 6748477542 residues

Word size : 0

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:\*  
1: em\_estbda:\*  
2: em\_estbhm:\*  
3: em\_estln:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hlc:\*  
9: gb\_estl:\*  
10: gb\_est2:\*  
11: gb\_hlc:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	49	4.2	383	10 BG938086	BG938086 1ab011a08
2	49	4.2	430	10 BG937697	BG937697 1ab005e01
3	49	4.2	472	10 BG938320	BG938320 1ab015e12
4	43	3.7	399	10 BG937723	BG937723 1ab005d06
5	43	2.0	325	9 AV532492	AV532492 AV532492
6	23	2.0	339	10 T20903	T20903 2911 Lambda
7	23	2.0	380	9 AV538262	AV538262 AV538262
8	23	2.0	391	9 AA969042	AA969042 op43d07.s
9	23	2.0	404	9 AA411567	AA411567 zv22901.s
10	23	2.0	408	9 AV518232	AV518232 AV518232
11	23	2.0	411	9 AV559941	AV559941 AV559941
12	23	2.0	438	9 AA411566	AA411566 zv22901.r
13	23	2.0	483	9 AV536021	AV536021 AV536021
14	23	2.0	505	9 A1479358	A1479358 tm27e07.x
15	23	2.0	511	9 AV518638	AV518638 AV518638
16	23	2.0	532	9 AV548208	AV548208 AV548208
17	23	2.0	549	9 AV557282	AV557282 AV557282

C 18	23	2.0	553	9 AV555646	AV555646 AV555646
C 19	23	2.0	565	9 AV532027	AV532027 AV532027
C 20	23	2.0	574	9 AV542531	AV542531 AV542531
C 21	23	2.0	581	9 AV439574	AV439574 AV439574
C 22	23	2.0	581	9 AV542248	AV542248 AV542248
C 23	23	2.0	640	9 AV537387	AV537387 AV537387
C 24	23	2.0	710	9 BE037991	BE037991 AA07D08.A
C 25	22	1.9	153	10 BE671871	BE671871 7448f09.x
C 26	22	1.9	341	12 AQ084497	AQ084497 HS_2252_A
C 27	21	1.8	622	9 AV734647	AV734647 AV734647
C 28	21	1.8	707	9 AL668938	AL668938 AL668938
C 29	21	1.8	842	12 AQ896484	AQ896484 HS_3134_A
C 30	21	1.8	1035	12 AG151623	AG151623 Pan trogl
C 31	20	1.7	257	12 BH093689	BH093689 Rpci-24-3
C 32	20	1.7	405	9 AL667069	AL667069 AL667069
C 33	20	1.7	531	10 BE353509	BE353509 EST353706
C 34	20	1.7	638	9 AV875590	AV875590 AV875590
C 35	20	1.7	807	9 AL664431	AL664431 AL664431
C 36	20	1.7	1101	12 CNS050HC	AL315561 Tetradon
C 37	19	1.6	211	9 AW754301	AW754301 CM0-CT033
C 38	19	1.6	235	10 N27368	N27368 yw52h07.s1
C 39	19	1.6	253	9 AV856918	AV856918 AV856918
C 40	19	1.6	284	9 AW749196	AW749196 MRA-BT035
C 41	19	1.6	298	9 AW754326	AW754326 CM1-CT033
C 42	19	1.6	301	9 AW754327	AW754327 CM1-CT033
C 43	19	1.6	308	9 AT282510	AT282510 gv04e03.x
C 44	19	1.6	311	9 AM858312	AM858312 CM0-CT033
C 45	19	1.6	312	9 AM858289	AM858289 CM0-CT033

#### ALIGNMENTS

RESULT 1  
BG938086 383 bp mRNA linear EST 11-JUN-2001  
LOCUS 1ab011a08 Bovine Abomasum cDNA Library Bos taurus cDNA 5', mRNA  
DEFINITION  
ACCESSION BG938086  
VERSION BG938086.1 GI:14337458  
KEYWORDS  
SOURCE  
ORGANISM  
Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovidae; Bovinae; Bos.  
1 (bases 1 to 383)  
Moore,S.S., Hansen,C., Li,C., Fu,A., Meng,Y. and Li,G.  
cDNA's from bovine abomasum tissue  
Unpublished (2001)  
Contact: Dr. Stephen Moore  
Beef Genomics Laboratory  
Dept of AFNS, University of Alberta  
410 Agri/For, Dept of AFNS, U of A, Edmonton, AB, T6G 2P5, Canada  
Tel: 780 492 0169  
Fax: 780 492 4265  
Email: smoores@afns.ualberta.ca  
The sequence best matches gb:BPJ19786 (Bos primigenius prochimostin  
mRNA, complete cds) in main database at high score of 735.0 and  
E-value of 0.0  
PCR Primers  
FORWARD: M13 Forward  
BACKWARD: M13 Reverse  
Seq primer: T3 primer  
High quality sequence stop: 383  
POLYA-No.

FEATURES  
source location/Qualifiers  
1..383  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/clone\_lib="Bovine Abomasum cDNA Library"  
/sex="Two males and one female mixed"  
/tissue\_type="Gastrointestinal tissue (GIT)"



AUTHORS Moore,S.S., Hansen,C., Li,C., Fu,A., Meng,Y. and Li,G.  
TITLE CDNA's from bovine abomasum tissue  
JOURNAL Unpublished (2001)  
COMMENT Contact: Dr. Stephen Moore  
. Beef Genomics Laboratory  
Dept of AFNS, University of Alberta  
410 Agri/For, Dept of AFNS, U of A, Edmonton, AB, T6G 2P5, Canada  
Tel: 780 492 0169  
Fax: 780 492 4265  
Email: smoores@afns.ualberta.ca

The sequence best matches gb:BP019786 (Bos primigenius prothymosin mRNA, complete cds) in main database at high score of 767.0 and E-value of 0.0

## PCR Primers

FORWARD: M13 Forward  
BACKWARD: M13 Reverse  
Seq primer: 73 primer  
High quality sequence stop: 399  
POLYA-No.

## FEATURES

## Source

Location/Qualifiers  
1..399  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/clone\_lib="Bovine Abomasum CDNA Library"  
/sex="Two males and one female mixed"  
/tissue\_type="Gastrointestinal tissue (GIT)"  
/cell\_type="Epithelial"  
/dev\_stage="Young adult"  
/lab\_host="XLI-BlueMRF"-strain"  
/note="Organ: Abomasum; Vector: Uni-2ZAPYR; Site\_1: Ecor  
I; Site\_2: Xho I"  
BASE COUNT 89 a 121 c 112 g 77 t  
ORIGIN

Query Match 3.7%; Score 43; DB 10; Length 399;  
Best Local Similarity 100.0%; Pred. No. 5.3e-11;  
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 662 cgggtacatgacaggaatgcccagagagatcctcagct 704  
|||||  
Db 1 CGGTTACATGACAGGAATGCGCAGAGACATGCTCAGCT 43

## RESULT 5

## AV532492/c

LOCUS AV532492 Arabidopsis thaliana flower buds Columbia Arabidopsis  
DEFINITION thaliana CDNA clone FB043b03f 3', mRNA sequence.  
ACCESSION AV532492  
VERSION AV532492.1 GI:8692775  
KEYWORDS EST.

## SOURCE

## ORGANISM

thale cress.  
Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi

REFERENCE  
AUTHORS Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.  
TITLE A large scale analysis of cDNA in Arabidopsis thaliana: Generation  
of 12,028 non-redundant expressed sequence tags from normalized and  
size-selected cDNA libraries

JOURNAL  
MEDLINE DNA Res. 7, 175-180 (2000)  
20363093

## COMMENT

Contact: Erika Asamizu  
The First Laboratory for Plant Gene Research  
Kazusa DNA Research Institute  
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan  
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/  
Location/Qualifiers

## FEATURES

## Source

1..325  
/organism="Arabidopsis thaliana"  
/strain="Columbia"

/db\_xref="taxon:3702"  
/clone\_lib="FB043b03f"  
/clone\_lib="Arabidopsis thaliana flower buds Columbia"  
/tissue\_type="flower buds"  
/note="Vector: pBluescriptII SK-; Site\_1: EcorI; Site\_2:  
XhoI"

BASE COUNT 106 a 92 c 59 g 68 t  
ORIGIN

Query Match 2.0%; Score 23; DB 9; Length 325;  
Best Local Similarity 100.0%; Pred. No. 1.1;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1087 tggatcttgagagatgtgtcat 1109  
|||||  
Db 209 TGGATCTTGGAGATGTGTTCAT 187

## RESULT 6

## T20903

LOCUS T20903 339 bp mRNA linear EST 07-JAN-1998  
DEFINITION 2911 Lambda-PRL2 Arabidopsis thaliana CDNA clone 89F977, mRNA  
sequence.  
ACCESSION T20903  
VERSION T20903.1 GI:2756823  
KEYWORDS EST.

## SOURCE

## ORGANISM

thale cress.  
Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi

REFERENCE  
AUTHORS Newman,T., deBrujn,F.J., Green,P., Keegstra,K., Kende,H., McIntosh  
L., Ohlrogge,J., Raikhel,N., Somerville,S., Thomashow,M., Retzel  
E. and Somerville,C.  
Genes galore: a summary of methods for accessing results from  
large-scale partial sequencing of anonymous Arabidopsis cDNA clones  
Plant Physiol. 106, 1241-1255 (1994)

JOURNAL  
MEDLINE 9518729  
COMMENT On Jan 7, 1998 this sequence version replaced gi:502344.  
Contact: Thomas Newman  
MSU-DOE Plant Research Laboratory  
Michigan State University  
MSU-DOE-PRL, Michigan State University, Plant Biology Bldg., E.  
Lansing, MI  
Tel: 517-353-0854  
Fax: 517-353-9168  
Email: 223131cn@dm.cl.msu.edu  
Seq primer: 77.

## FEATURES

## Source

Location/Qualifiers  
1..339  
/organism="Arabidopsis thaliana"  
/strain="var Columbia"  
/db\_xref="taxon:3702"  
/clone="89F977"  
/clone\_lib="Lambda-PRL2"

/note="Vector: lambda Zip-Lox; Site\_1: Sal; Site\_2: Not;  
Lambda PRL2 is a cDNA library derived from equal  
quantities of 4 pools of mRNA. The mRNA sources were 1) 7  
day germinated etiolated seedlings; 2) tissue culture  
grown roots; 3) staged plants half with 24 hour light  
cycle, half on 16 hr light, 8 hour dark- rosettes; 4)  
same plants as 3 but aerial tissue (stems, flowers and  
siliques). The vector is BRL's lambda Zip-Lox. The cDNA  
inserts were directionally cloned with Sal-Not arms using  
oligo dt primed cDNA."

BASE COUNT 67 a 61 c 92 g 106 t 13 others  
ORIGIN

Query Match 2.0%; Score 23; DB 10; Length 339;  
Best Local Similarity 100.0%; Pred. No. 1.1;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1087 tggatcttggagatgtgtcat 1109  
 |||  
 Db 82 TGGATCTTGGGAGATGTGTTCA 104

RESULT 7  
 AV538262 380 bp mRNA linear EST 06-SEP-2000  
 AV538262 Arabidopsis thaliana roots Columbia Arabidopsis thaliana  
 LOCUS CDNA clone R2113a03f 3', mRNA sequence.

ACCESSION AV538262  
 VERSION AV538262.1 GI:8698545

KEYWORDS EST.  
 SOURCE thale cress.

ORGANISM Arabidopsis thaliana

REFERENCE Arabidopsis thaliana: Streptophyta; Embryophyta; Tracheophyta; Eukaryota; Viridiplantae; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

A large scale analysis of cDNA in Arabidopsis thaliana: Generation of 12,028 non-redundant expressed sequence tags from normalized and size-selected cDNA libraries

Asamizu, E., Nakamura, Y., Sato, S., and Tabata, S.  
 1 (bases 1 to 380)  
 TITLE DNA Res. 7, 175-180 (2000)

JOURNAL 20363093

COMMENT Contact: Erika Asamizu  
 The First Laboratory for Plant Gene Research  
 Kazusa DNA Research Institute  
 Yana 1532-3, Kisarazu, Chiba 292-0812, Japan  
 Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.  
 Location/Qualifiers

FEATURES  
 source  
 1..380  
 /organism="Arabidopsis thaliana"  
 /strain="Columbia"  
 /db\_xref="taxon:3702"  
 /clone="R2113a03f"  
 /clone\_lib="Arabidopsis thaliana roots Columbia"  
 /tissue="roots"  
 /note="Vector: pBluescriptII SK-; Site\_1: EcoRI; Site\_2: XhoI"

BASE COUNT 730 a 96 c 65 g 89 t

Query Match 2.0%; Score 23; DB 9; Length 380;  
 Best Local Similarity 100.0%; Pred. No. 1.1;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1087 tggatcttggagatgtgtcat 1109  
 |||  
 Db 332 TGGATCTTGGGAGATGTGTTCA 310

RESULT 8  
 AA969042 391 bp mRNA linear EST 07-JUL-1998  
 LOCUS op3d07.sl Soares.NFL.T.GBC.SI Homo sapiens CDNA clone  
 DEFINITION IMAGE:1579597 3' similar to TR:Q27951 Q27951 PROCHYMOSIN. ; mRNA  
 sequence.

ACCESSION AA969042  
 VERSION AA969042.1 GI:3144222

KEYWORDS EST.  
 SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

1 (bases 1 to 391)  
 NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)

## COMMENT

Contact: Robert Strausberg, Ph.D.  
 Email: cgaps@rmail.nih.gov  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.lnl.gov) for further information.  
 Insert Length: 619 Std Error: 0.00  
 Seq primer: -40m13 fwd. ET from Amersham  
 High quality sequence stop: 378.  
 Location/Qualifiers

## FEATURES

## source

1..391  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:1579597"  
 /clone\_lib="Soares.NFL.T.GBC.SI"  
 /lab\_host="DH10B"  
 /note="Organ: pooled; Vector: pUT3D-Pac (Pharmacia) with  
 a modified polylinker. Site\_1: Not I; Site\_2: Eco RI.  
 Equal amounts of plasmid DNA from three normalized  
 libraries (fetal lung NbHL19W, testis NHT, and B-cell  
 NCI-CCGAP GCBI) were mixed, and ss circles were made in  
 vitro. Following HAP purification, this DNA was used as  
 tracer in a subtractive hybridization reaction. The driver  
 was PCR-amplified cDNAs from pools of 5,000 clones made  
 from the same 3 libraries. The pools consisted of  
 1.M.A.G.E. clones 297480-302087, 682632-687239,  
 726408-728711, and 729096-731399. Subtraction by Bento  
 Soares and M. Fatima Bonaldo.  
 BASE COUNT 82 a 101 c 115 g 93 t

Query Match 2.0%; Score 23; DB 9; Length 391;  
 Best Local Similarity 100.0%; Pred. No. 1.2;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1015 cccctcgctataccagcagga 1037  
 |||  
 Db 256 CCTCCGCTATACACAGCAGGA 234

RESULT 9  
 AA411567 404 bp mRNA linear EST 17-MAY-1997  
 LOCUS zv22901.sl Soares.NHMPU.SI Homo sapiens CDNA clone IMAGE:75416 3'  
 DEFINITION similar to SW:CHYM\_SHEEP P18276 PROCHYMOSIN PRECURSOR, B-FORM ;  
 mRNA sequence.

ACCESSION AA411567  
 VERSION AA411567.1 GI:2069151

KEYWORDS EST.  
 SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

1 (bases 1 to 404)  
 Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,  
 Kucaba, T., Lacy, M., Lee, N., Lennon, G., Maria, M., Martin, D., Moore, B.,  
 Schellenberg, K., Steptoe, M., Tan, F., Theising, B., White, Y., Wyllie,  
 T., Waterston, R., and Wilson, R.

WashU-Merck EST Project 1997  
 Unpublished (1997)  
 CONTACT: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810

Email: est@watson.wustl.edu  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.lnl.gov) for further information.  
 Possible reversed clone: similarity on wrong strand  
 Seq primer: -41m13 fwd. ET from Amersham  
 High quality sequence stop: 374.

## FEATURES

## source

1..404  
 /organism="Homo sapiens"



```

/db_xref="GDB:5977315"
/db_xref="taxon:9606"
/clone_image="754416"
/clone_lib="Soares.NbHMPU_S1"
/tissue_type="Pooled human melanocyte, fetal heart, and
pregnant uterus"
/lab_host="DH10B"
/notes="Organ: mixed (see below); Vector: pT7T3D-Pac
(Pharmacia) with a modified polylinker; Site_1: Not 1;
Site_2: Eco RI; Equal amounts of plasmid DNA from three
normalized libraries (melanocyte 2NbHMP, pregnant uterus
NbHMP, and fetal heart NbHMP19W) were mixed, and ss circles
were made in vitro. Following HAP purification, this DNA
was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from pools of
5,000 clones made from the same 3 libraries. The pools
consisted of I.M.A.G.E. clones 260232-265223,
340488-345479, and 484488-489479."

BASE COUNT      88 a      105 c      116 g      95 t
ORIGIN

Query Match      2.0%; Score 23; DB 9; Length 404;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1015 cccctcgctaccaccagga 1037
|||||
Db 257 CCTCCGCTTACGACGACGGA 235

RESULT 10
AV518232      408 bp      mRNA      linear      EST 06-SEP-2000
LOCUS
DEFINITION
AV518232 Arabidopsis thaliana aboveground organs two to six-week
old Arabidopsis thaliana cDNA clone APD13d01F 3', mRNA sequence.
ACCESSION
AV518232
VERSION
AV518232.1 GI:86777559
KEYWORDS
SOURCE
thale cress.
ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
1 (bases 1 to 408)
A large scale analysis of cDNA in Arabidopsis thaliana: Generation
of 12,028 non-redundant expressed sequence tags from normalized and
size-selected cDNA libraries
DNA Res. 7, 175-180 (2000)
20363093
COMMENT
Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
Location/Qualifiers
1. 408
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone_image="APD13d01F"
/clone_lib="Arabidopsis thaliana aboveground organs two to
six-week old"
/tissue_type="aboveground organs"
/dev_stage="two to six-week old"
/notes="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
XhoI"

BASE COUNT      135 a      112 c      69 g      92 t
ORIGIN

Query Match      2.0%; Score 23; DB 9; Length 408;

```

```

Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1087 tggatctggagatgtgtcat 1109
|||||
Db 285 TGGATCTTGGGAGATGTGTCA 263

RESULT 11
AV559941/c      411 bp      mRNA      linear      EST 07-SEP-2000
LOCUS
DEFINITION
AV559941 Arabidopsis thaliana green siliques Columbia Arabidopsis
thaliana cDNA clone SQ126c01F 3', mRNA sequence.
ACCESSION
AV559941
VERSION
AV559941.1 GI:8731367
KEYWORDS
SOURCE
EST.
thale cress.
ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
1 (bases 1 to 411)
A large scale analysis of cDNA in Arabidopsis thaliana: Generation
of 12,028 non-redundant expressed sequence tags from normalized and
size-selected cDNA libraries
DNA Res. 7, 175-180 (2000)
20363093
COMMENT
Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
Location/Qualifiers
1. 411
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone_image="SQ126c01F"
/clone_lib="Arabidopsis thaliana green siliques Columbia"
/tissue_type="green siliques"
/notes="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
XhoI"

BASE COUNT      136 a      109 c      70 g      96 t
ORIGIN

Query Match      2.0%; Score 23; DB 9; Length 411;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1087 tggatctggagatgtgtcat 1109
|||||
Db 332 TGGATCTTGGGAGATGTGTCA 310

RESULT 12
AA411566      438 bp      mRNA      linear      EST 17-MAY-1997
LOCUS
DEFINITION
zv22901.f1 Soares.NbHMPU_S1 Homo sapiens cDNA IMAGE:754416 5'
similar to TR:G457097 G457097 CHYMOSIN C. [1] ;, mRNA sequence.
ACCESSION
AA411566
VERSION
AA411566.1 GI:2069150
KEYWORDS
SOURCE
EST.
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 438)
Hillier, L., Allen, M., Bowles, L., Dubuque, T., Giesel, G., Jost, S.,
Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B.,
Schellenberg, K., Stepien, M., Tan, F., Theisling, B., White, Y., Wyllie,
T., Waterston, R. and Wilson, R.

```

TITLE Washu-Merck EST Project 1997  
JOURNAL Unpublished (1997)  
COMMENT Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: -28m13 rev2 ET from Amer sham.

FEATURES  
source

1. 438  
/organism="Homo sapiens"  
/db\_xref="GDB:5977315"  
/db\_xref="taxon:9606"  
/clone\_lib="IMAGE:754416"  
/clone\_lib="Soares\_NhMMPu\_S1"  
/tissue\_type="Pooled human melanocyte, fetal heart, and  
pregnant uterus"  
/lab\_host="DH10B"  
/note="Organ: mixed (see below); Vector: pT73D-Pac  
(Pharmacia) with a modified polylinker. Site\_1: Not I;  
Site\_2: Eco RI; Equal amounts of plasmid DNA from three  
normalized libraries (melanocyte 2NDH, pregnant uterus  
NbHPV, and fetal heart NbH19W) were mixed, and ss circles  
were made in vitro. Following HAP purification, this DNA  
was used as tracer in a subtractive hybridization  
reaction. The driver was PCR-amplified cDNAs from pools of  
5,000 clones made from the same 3 libraries. The pools  
consisted of 1 M.A.G.E. clones 260232-265223,  
340488-345479, and 484488-489479."

BASE COUNT 103 a 124 c 117 g 94 t  
ORIGIN  
Query Match 2.0%; Score 23; DB 9; Length 438;  
Best Local Similarity 100.0%; Pred. No. 1.2;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Y 1015 cccctcgctataccagcagga 1037  
|||||  
Db 187 CCCCGCGCTATACCACGACGAGA 209

RESULT 13  
AV536021/c 483 bp mRNA linear EST 06-SEP-2000  
LOCUS AV536021 Arabidopsis thaliana flower buds Columbia Arabidopsis  
DEFINITION thaliana cDNA clone FB106C09F 3', mRNA sequence.  
ACCESSION AV536021  
VERSION AV536021.1 GI:8696304  
KEYWORDS EST.  
SOURCE thale cress.  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
1 (bases 1 to 483)  
Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.  
A large scale analysis of cDNA in Arabidopsis thaliana: Generation  
of 12,028 non-redundant expressed sequence tags from normalized and  
size-selected cDNA libraries  
DNA Res. 7, 175-180 (2000)  
20363093  
JOURNAL MEDLINE  
COMMENT Contact: Erika Asamizu  
The First Laboratory for Plant Gene Research  
Kazusa DNA Research Institute  
Yena 1532-3, Kisarazu, Chiba 292-0812, Japan  
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.  
Location/Qualifiers  
FEATURES  
source  
1. 483  
/organism="Arabidopsis thaliana"

/strain="Columbia"  
/db\_xref="taxon:3702"  
/clone\_lib="FB106C09F"  
/clone\_lib="Arabidopsis thaliana flower buds Columbia"  
/tissue\_type="flower buds"  
/note="Vector: pBluescriptII SK-; Site\_1: EcoRI; Site\_2:  
XhoI"

BASE COUNT 161 a 123 c 85 g 114 t  
ORIGIN  
Query Match 2.0%; Score 23; DB 9; Length 483;  
Best Local Similarity 100.0%; Pred. No. 1.2;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Y 1087 tggatctggagagatgttcat 1109  
|||||  
Db 338 TGGATCTTGGAGATGTGTTTCAT 316

RESULT 14  
A1479358/c 505 bp mRNA linear EST 14-APR-1999  
LOCUS A1479358  
DEFINITION tm27607.x1 Soares\_NFL\_T\_GBC\_S1 Homo sapiens cDNA clone  
IMAGE:2157828 3' similar to TR:028950 Q28950 PREPROCHYMOSIN  
PRECURSOR ; mRNA sequence.  
ACCESSION A1479358  
VERSION A1479358.1 GI:4372526  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 505)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index

JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: c9aps-r@email.nih.gov  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert Length: 608 Std Error: 0.00  
Seq primer: -400P from Gibco  
High quality sequence stop: 367.  
Location/Qualifiers

source  
1. 505  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="IMAGE:2157828"  
/clone\_lib="Soares\_NFL\_T\_GBC\_S1"  
/lab\_host="DH10B"  
/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with  
a modified polylinker. Site\_1: Not I; Site\_2: Eco RI;  
Equal amounts of plasmid DNA from three normalized  
libraries (fetal lung NbH19W, testis NHT, and B-cell  
NCL1CGAP.GCBI) were mixed, and ss circles were made in  
vitro. Following HAP purification, this DNA was used as  
tracer in a subtractive hybridization reaction. The driver  
was PCR-amplified cDNAs from pools of 5,000 clones made  
from the same 3 libraries. The pools consisted of  
1 M.A.G.E. clones 297480-302087, 682632-687239,  
726408-728711, and 729096-731399. Subtraction by Bento  
Soares and M. Fatima Bonaldo.  
BASE COUNT 106 a 138 c 141 g 119 t 1 others  
ORIGIN

Query Match 2.0%; Score 23; DB 9; Length 505;  
Best Local Similarity 100.0%; Pred. No. 1.2;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Y 1015 cccctcgctataccagcagga 1037

Db 271 CCCGCCCTATACCAGCAGGA 249

RESULT 15

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

FEATURES

source

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches

QY 1087

Db 274

Search completed: July 31, 2002, 23:07:30

job time: 17275 sec

511 bp mRNA linear EST 06-SEP-2000

AV518638 Arabidopsis thaliana aboveground organs two to six-week

old Arabidopsis thaliana cDNA clone APD34e06f 3', mRNA sequence.

AV518638

AV518638.1 GI:8678165

EST.

thale cress.

Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosidae; eustosids II; Brassicales; Brassicaceae; Arabidopsis.

1 (bases 1 to 511)

Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.

A large scale analysis of cDNA in Arabidopsis thaliana: Generation

of 12,028 non-redundant expressed sequence tags from normalized and

size-selected cDNA libraries

DNA Res. 7, 175-180 (2000)

20363093

Contact: Erika Asamizu

The First Laboratory for Plant Gene Research

Kazusa DNA Research Institute

Yana 1532-3, Kisarazu, Chiba 292-0812, Japan

Email: asamizukazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

Location/Qualifiers

1. 511

/organism="Arabidopsis thaliana"

/strain="Columbia"

/db\_xref="taxon:3702"

/clone="APD34e06f"

/clone\_lib="Arabidopsis thaliana aboveground organs two to

six-week old"

/tissue\_type="aboveground organs"

/dev\_stage="two to six-week old"

/note="Vector: pBluescriptII SK-; Site\_1: EcoRI; Site\_2:

XhoI"

156 a 135 c 102 g 118 t

2.0%; Score 23; DB 9; Length 511;

Best Local Similarity 100.0%; Pred. No. 1.2;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1087

Db 274

Search completed: July 31, 2002, 23:07:30

job time: 17275 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 1, 2002, 01:41:47 ; Search time 9238.6 Seconds  
(without alignments)  
8963.079 Million cell updates/sec

Title: US-09-643-755B-3  
Perfect score: 3957  
Sequence: 1 ctgcaggaattcatgtact.....acctcaccactaagtacc 3957

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 1797656 seqs, 10463268293 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_inv:\*  
32: em\_htg\_other:\*  
33: em\_htgo\_inv:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No. Query Match Length DB ID Description

1	3957	100.0	3957	6	AX088021	AX088021 Sequence
2	1551	39.2	3502	6	PVBCSP	J01263 Phaseolus v
3	1173	29.6	1173	6	AX088019	AX088019 Sequence
4	1047	26.5	1415	6	AX343913	AX343913 Sequence
5	887	22.4	2970	6	AX252300	AX252300 Sequence
6	181	4.6	1990	6	E00452	E00452 DNA sequenc
7	181	4.6	2090	6	A06496	A06496 Artificial
8	169	4.3	4764	6	PVAPHASE	X52626 Phaseolus v
9	133	3.4	1475	6	A11822	A11822 Artificial
10	133	3.4	1475	6	PVPHASBR	X03004 Phaseolus v
11	95	2.4	207	8	ZMEINI	X06175 Zea mays/P.
12	90	2.3	1535	8	PVPHASAR	X02960 Phaseolus v
13	77	1.9	1454	8	U01131	U01131 Phaseolus v
14	75	1.9	348	8	ZMEINI2	X06176 Maize chime
15	63	1.6	1478	8	U01121	U01121 Phaseolus l
16	57	1.4	1430	8	U01132	U01132 Phaseolus v
17	56	1.4	1269	4	BOVCHYMOC	J00004 Bovine chym
18	49	1.2	1098	6	E00079	E00079 CDNA sequen
19	49	1.2	1098	6	E02341	E02341 CDNA sequen
20	49	1.2	1101	4	AF177290	AF177290 Bubalus a
21	49	1.2	1115	12	SYNPROCA	M22593 Synthetic b
22	49	1.2	1175	6	I04058	I04058 Sequence 5
23	49	1.2	1240	6	AR002347	AR002347 Sequence
24	49	1.2	1275	4	BOVCHYMOA	J00002 bovine chym
25	49	1.2	1289	6	E00144	E00144 CDNA encodi
26	49	1.2	1291	6	A15836	A15836 chymosin ge
27	49	1.2	1305	4	BOVCHYMOB	J00003 bovine chym
28	49	1.2	1460	6	E00042	E00042 DNA coding
29	49	1.2	1460	6	E00285	E00285 CDNA encodi
30	49	1.2	2726	6	I08097	I08097 Sequence 5
31	49	1.2	2733	6	AR073077	AR073077 Sequence
32	49	1.2	2982	6	I08098	I08098 Sequence 1
33	44	1.1	1094	4	BP019786	U19786 Bos primlge
34	38	1.0	1290	6	E00075	E00075 CDNA encodi
35	38	1.0	1291	6	A15633	A15633 preprochymo
36	38	1.0	1311	6	E00108	E00108 DNA coding
37	37	0.9	119	4	BOVCHYMO8	M14076 Bovine chym
38	37	0.9	1404	8	U01122	U01122 Phaseolus l
39	35	0.9	39	4	BOVCHYMO2	M14070 Bovine chym
40	35	0.9	637	6	E05472	E05472 DNA sequenc
41	33	0.8	179	6	I04059	I04059 Sequence 6
42	33	0.8	1172	4	CDR131677	A0131677 Camelus d
43	32	0.8	335	4	BOVCHYMO9	M14077 Bovine chym
44	32	0.8	1292	4	OAPPCHY	X53037 Lamb mRNA f
45	30	0.8	134	4	BOVCHYMO6	M14074 Bovine chym

ALIGNMENTS

RESULT 1  
AX088021  
LOCUS AX088021  
DEFINITION Sequence 3 from Patent WO0114571.  
ACCESSION AX088021  
VERSION AX088021.1 GI:13396949  
KEYWORDS  
SOURCE  
ORGANISM  
synthetic construct.  
synthetic construct.  
artificial sequence.  
REFERENCE  
AUTHORS 1 (bases 1 to 3957)  
van Rooijen,G., Keon,R.G., Boothe,J. and Shen,Y.  
TITLE Commercial production of chymosin in plants  
JOURNAL Patent: WO 0114571-A 3 01-MAR-2001;  
Sembiosys Genetics Inc. (CA)  
FEATURES  
source Location/Qualifiers  
1..3957  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
1554..2726  
/note="Figure 2"  
/codon\_start=1  
/transl\_table=11



Db 1921 AAAATTGATCCGAGAAAGCGTCACCTTCGAACTTAGGCAAAACCTTGTCTATAC 1980  
Oy 1981 actacggtacaggtacagatgtaagaagaaacttaagctatgtacacgtatctccaca 2040  
Db 1981 ACTACGGTACAGGATGATGTAAGAAATCTTAGCTATGATACCGTCACTGTCCAAACA 2040  
Oy 2041 ttgtgacatccaacagacagtaagacttagcacccaagaaccagtgatgtcttcacct 2100  
Db 2041 TTGTGACATTCACACAGACAGTAGACTTAGCACCAAGAACAGGTATGTCTTCACCT 2100  
Oy 2101 atgcagaatcgcattgcatcccttggtatgagcaaccacgcgtcgcgcagagtaactga 2160  
Db 2101 ATGAGAAATTCGATGCGATCCTTGTGTGATGCGATCCCATCCTCGCGCAAGACTGCA 2160  
Oy 2161 taactggtttgacaacatagatgaacgcgaaccctagagctcaagacttgcttcggttt 2220  
Db 2161 TACCTGTTGTAACAATGATGTAACCGACACCTAGTAGCTCAAGACTTGTTCTCGGTTT 2220  
Oy 2221 acatgagcagaatgagccagagagacatgctcaacgcttgagctatgatccatccact 2280  
Db 2221 ACATGAGACAGGAATGGCGAGAGACATGCTCAGCGTTGAGCTATGTATCATTCTACT 2280  
Oy 2281 acacaggaatcctctcacttgggttccagtcacgtgtgcagcagtaactggcaatcactg 2340  
Db 2281 ACACAGGATCTCTCACTGGGTTCCAGTCACTGTCAGACAGTACTGGCAATTCCTG 2340  
Oy 2341 acaggtgcacacatcagaggtggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2400  
Db 2341 ACAGTGCACCATCAGCGGT 2400  
Oy 2401 ataaccgttaacgttccaaagctgtgcagcctagcagcagcattctcaacatcagaaccta 2460  
Db 2401 ATACCGGTAGTCCAACTGCTGCGACCTAGACACGACATTTCTCAACTTAGAGACGTA 2460  
Oy 2461 ttggaagcacacagaaacagtaacgttgaagttgacatagattgagcaacacttaactga 2520  
Db 2461 TTGGAGCCACACAGAACACAGTACGTTGATGATGATGATGATGATGATGATGATGAT 2520  
Oy 2521 tgcctaaagtgtgtcttggatagatcaacggcgaagatgtacccactgaacccctccgctata 2580  
Db 2521 TGCTTACAGTGTCTTTGAGATCAACGGCAGATGATCCCTGACCCCTCCCTCTATA 2580  
Oy 2581 ccaagccagatcaaaaggtttctgcacagtgatccagagtgagtaacatccacagaat 2640  
Db 2581 CCAAGCCAGATCAAGGGTTCTGCGACAGTGTATCCAGAGTGAACCATTTCCAGAAAT 2640  
Oy 2641 ggaatctggagagatgtgtcaatcgtgagtaactaacgcttctgacagggccaacaac 2700  
Db 2641 GGATCTTGGGAGATGTGTCTATCTGATGATCTACACGCTTTGACAGGGCCAAACAC 2700  
Oy 2701 tcgtttggctcagctaaagaacatctgaagcttaataagtaagtaactaaatgtatggt 2760  
Db 2701 TCGTTGGGCTAGCTAAAGCAATCTGAAGCTTAATTAAGTAACTAAATTAAGTAACTAGT 2760  
Oy 2761 gtaagaagcctaagagacatgagatactgtatccgacatgtaacagtaataactga 2820  
Db 2761 GTAAGACCTCATGAGAGCATGGAATGTGTATCCGACATGTAACAGTATTAATTAAGTGA 2820  
Oy 2821 gctcatctcaacttctctatgaataaacaagaagtgtaagataataaactctatc 2880  
Db 2821 GCTCATCTCACTCTCTATGATTAACAAGATGTATGATATTAACACTCTATC 2880  
Oy 2881 taagacactatgttctatagataaatttcctctatatataatacatcgaatggtg 2940  
Db 2881 TAAGACCTTATGTTTATGATTAATTTCTCTTATTAATTAATCAATCTGATGATG 2940  
Oy 2941 acgagctatggaatgctcacaatagtaacaaacaacatgtgtactataagaactctcaaa 3000  
Db 2941 ACGGCTTATGGAATGCTTCAAAATGTATCAAAAACAAATGTGTACTATTAAGCTTTCTAAA 3000  
Oy 3001 caatttaacttaagcatgttgaagagacataagtgcttaagaagacataaactataa 3060

Db 3001 CAATTTCACTTTAGCATTTGAAAGAGACATTAAGTGTAAAGAACATTAATTAATA 3060  
Oy 3061 tggagaaggttgcctccattatataataattaccactatgataataatga 3120  
Db 3061 TGGAGAAAGTTGTCTCATTTATATTAATTAATTAATTAATTAATTAATTAATTAATGA 3120  
Oy 3121 tgttaagagacataacatataaagagagaagtttlatccatllatataatataac 3180  
Db 3121 TGTTAAGAGACATTAACATTAATTAAGAGAAAGTTGTATCCATTTATATTAATTAATAC 3180  
Oy 3181 taaccattatataataactatccactatlttaatgcttataaagttgatccatg 3240  
Db 3181 TACCATTTATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3240  
Oy 3241 atattcctaattttagttgataatgataagaaggtaactattggaacctctactc 3300  
Db 3241 ATATTCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3300  
Oy 3301 tglataaagttgatacctcctaaagtgggtcctatataattatctctcctacagat 3360  
Db 3301 TGTATTAAGGTGGATCATCTCTTAAGGGGCTATTAATTTATTTCTCTTACAGAT 3360  
Oy 3361 aaaaaaaatlataggttgggttgataaaatlatgaagatltaaaataataataa 3420  
Db 3361 AAAAAAAATTAATGAGTTGGTTGATTAATTAATTAATTAATTAATTAATTAATTAATTA 3420  
Oy 3421 ataataacataatataatgataataaattatataataataataataataataa 3480  
Db 3421 ATAAATTAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3480  
Oy 3481 agtaataatgtcataaaccataacatcgttttagcccttgtagcagcctcactaatt 3540  
Db 3481 AGTAATATTTTCTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3540  
Oy 3541 taacagagatgaacataatgtgacttttggtatalttaacaataatlatataactat 3600  
Db 3541 TAAACGAGAGTAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3600  
Oy 3601 atgaatattttttttttatgtgcaagaaataaataattgaggaggcaatggt 3660  
Db 3601 ATGAATATTTTATTTTATGCAAGGAATAATTAATTAATTAATTAATTAATTAATTAAT 3660  
Oy 3661 gtcccaatcctctatacaaccacttccacaggaagtgtagtgcgggacacaaacaaac 3720  
Db 3661 GTCCCAATCCTTATTAACAACCAATCTCCACAGAGTGTAGTCCGGGACCAACAAAAAAC 3720  
Oy 3721 aggcagaggaaattttaaatttggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3780  
Db 3721 AGGCAAGGAATAATTTTAATTTGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3780  
Oy 3781 cactacacataacccttttagcagtagagcaatggttgaccgtgtgcttaactcttla 3840  
Db 3781 CACTACACATTAACCTTTTACAGATGAGCAATGCTGTACCGTGTGTGTGTGTGTGTGT 3840  
Oy 3841 ttctatttttctacagcaagaataaataaataaataaataaataaataaataaataaataa 3900  
Db 3841 TTTTATTTTATTAACAGAAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3900  
Oy 3901 aacctatatacaaaaccccaaaacaaagttctcctagaacccctacaactaaagtaacc 3957  
Db 3901 AACCTTATACAAAACCCCAAAAACAAAGTTCTTACGACCCCTACCACTAAGTAAAC 3957

## RESULT 2

PHVBCSP PHVBCSP 3502 bp DNA linear PLN 17-JUN-1998  
LOCUS Phaseolus vulgaris beta-type phaseolin storage protein gene,  
DEFINITION complete cds.  
ACCESSION J01263 M13758  
VERSION J01263.1 GI:3228361  
KEYWORDS  
SOURCE Phaseolus vulgaris.  
ORGANISM Phaseolus vulgaris

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Phaseolus

1 (bases 1371 to 2450; 2466 to 3294; 3322 to 3502)

Phaseolus

Slightom,J.L., Sun,S.S.M. and Hall,T.C.  
Complete nucleotide sequence of a French bean storage protein gene: phaseolin

Proc. Natl. Acad. Sci. U.S.A. 80, 1897-1901 (1983)

2 (bases 1 to 3502)

Doyle,J.J., Schuller,M.A., Godette,W.D., Zenger,Y., Beachy,R.N. and Slightom,J.L.  
The glycosylated seed storage proteins of Glycine max and Phaseolus vulgaris: Structural homologies of genes and proteins

J. Biol. Chem. 261, 9228-9238 (1986)

3 (bases 2921 to 3502)

Hall,T.  
Direct Submission  
Submitted (27-APR-1993) Biology, Texas A&M University, College Station, TX 77843-3155, USA

4 (bases 1 to 3502)

Hall,T.  
Direct Submission  
Submitted (16-JUN-1998) Biology, Texas A&M University, College Station, TX 77843-3155, USA

Nucleotide sequence updated by submitter  
On Jun 16, 1998 this sequence version replaced gi:169322.

REMARK  
COMMENT  
FEATURES  
SOURCE

1. 3502

/organism="Phaseolus vulgaris"  
/db\_xref="taxon:3885"  
/clone="AF-pvph3.8; pvph 177.4-E1.6"  
/clone\_1lb="Charon 24a clone 177.4"  
join(1548..1860,1933..2123,2212..2292,2417..2662,  
2791..3049,3153..3370)  
/gene="beta-psp"  
/number=1  
join(1548..1860,1933..2123,2212..2292,2417..2662,  
2791..3049,3153..3370)  
/gene="beta-psp"  
/codon\_start=1  
/product="beta-type phaseolin storage protein"  
/protein\_id="AAC23610.1"  
/db\_xref="GI:169323"

/translation="MMARVPLLLGLFLASISAFATSLREEESQNPFTFNSDN  
SNWTLFRKNOGHIRVLQRFDOQSKRLQNEIDYRLVERSKPETLLPQDADELLV  
RSGSALIVLVKPDREYFPLTSDNPIFSHQKIPAGTIFYLVNPPKEDRLIOLAM  
PVNPNQIHEFFLSTEAQOQYLQEFKSHLEAFNSFEEINRVLFEEOQEOGOE  
GIVINIDSEQIKELSKHAKSSRSKSKONTGNEGNLTERTDNLNVLISIEHE  
EGALFVPHYSKAIVILVNEGRAHVLNPKGNKLTPEESYRAELSKDQVVIPIAA  
YPAIKATSNVNTFTGFINANNNNRNILAKTQNVISITRDLGKDYLLTRTSGSD  
EVNKLINQSGSFVDAAHHQDEQKSGHOEQKSGKGAFFY"  
1548..1619  
/gene="beta-psp"  
join(1620..1860,1933..2123,2212..2292,2417..2662,  
2791..3049,3153..3367)  
/gene="beta-psp"  
/product="beta-type phaseolin storage protein"  
1861..1932  
/gene="beta-psp"  
/number=1  
1933..2123  
/gene="beta-psp"  
/number=2  
2124..2211  
/gene="beta-psp"  
/note="putative"  
/number=2  
2212..2292  
/gene="beta-psp"

sig\_peptide  
mat\_peptide  
intron  
exon  
intron  
exon

intron  
exon  
intron  
exon  
intron  
exon

/number=3  
2293..2416  
/gene="beta-psp"  
/note="putative"  
/number=3  
2417..2662  
/gene="beta-psp"  
/number=4  
2663..2790  
/gene="beta-psp"  
/note="putative"  
/number=4  
2791..3049  
/gene="beta-psp"  
/number=5  
3050..3152  
/gene="beta-psp"  
/note="putative"  
/number=5  
3153..>3370  
/gene="beta-psp"  
/number=6

BASE COUNT 1124 a 678 c 572 g 1128 t  
ORIGIN

Query Match 39.2%; Score 1551; DB 8; Length 3502;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1551; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 gaattcattgtactccagatcatcatlaagtgaaagtttggctctcgcggtggtt 66  
|||||  
Db 1 GAATTCATTGTACTCCAGTATCATTAATAGGAAAGTTTGCTGCTCGCGGTGTT 60  
QY 67 ttaccctattaaaggggtttccacctaataattcgtatcatcttacttact 126  
|||||  
Db 61 TTTACCTCTATTAAAGGGGTTTCCACCTAAATTCGTGATCATTCCTACTTACTT 120  
QY 127 gttaacttaattctcatatccttcttgtaaatcatcgcctccgacacgatatcc 186  
|||||  
Db 121 GTTACTTTAATTCTCATTAATCTTTGGTTGAATATACCGCTTCCGACACGATATCC 180  
QY 187 tacaatttattatcttgtaaacatttcaaacccgataaaatttlaagatccgct 246  
|||||  
Db 181 TACAATTTATTATTGTTAATCAATTTCAACCGCATAAATTTTGAAGTCCGCTCT 240  
QY 247 atctttaatgagctcaacatttcatatgtaataataataattacttaattagct 306  
|||||  
Db 241 ATCTTTAATGTAAGTCTTAACATTTTCATATTAATAATAATAATTTTACGCTT 300  
QY 307 gtagaagacataagatttattcttattcttctcatataaagtttaataataat 366  
|||||  
Db 301 GGTAGAAAGCATTAAGATTATTCTTATCTCTTCATATTAATGTTTAATATCAATAT 360  
QY 367 aaacaattcttaacctaaagaagatttcccatltatattttaaataatattca 426  
|||||  
Db 361 AACCAATCTTACCTTAAGAAGATTCCCATTTTATATTTAAAAATATTATTATCA 420  
QY 427 aatatlttcaacacgtaaatctcataataataagttgtttcaaaagtaataaatta 486  
|||||  
Db 421 AATATTTTTCACACGTAATATCTCATTAATTAAGTTGTTTCAAAAGTAATTAATTTA 480  
QY 487 actcataatttttatttgactgactgatttaagcaaacaccagtgacacaactagcat 546  
|||||  
Db 481 ACTCATATATTTTATTTCAGCTGATCTTAAAGCAACACCAAGTACACACTAGCCAT 540  
QY 547 tttttcttgaataaaaaaaccatcatcatgttattttttatataaagtaaaatt 606  
|||||  
Db 541 TTTTTCCTTGAATAAAAAATCAATATGATGTAATTTTATATACATTAATAATTTT 600  
QY 607 caccacaacatcatgtgtatattcttgaagaagatcatgttatagcaaatctcataat 666  
|||||  
Db 601 CACCAACAAATCATTTGGTATTTGTGAAGCAAGTATGATGTAACAAATTTATATTT 660



QY 667 ccacattgacacacggaagtaacatgacatcgtcttatacatgacagacacatctcta 726  
|||||  
Db 661 CCATTGACACCTACGAGAGTAAGATCTGCTTTTACATGCGAGACACATCTTCTTA 720  
QY 727 aagtaatttaataatgactataatcaagatttcaatalatacaataactaatattac 786  
|||||  
Db 721 AAGTAATTTTAATTAATGTTACTATATTCAGAAATTCATATATCAATAATATATAC 780  
QY 787 ttctaaataataatgataataataataatcttttataatttaatttaattgt 846  
|||||  
Db 781 TTCTAAATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 840  
QY 847 tgaattgtacatacttattatctactactactactactactactactactactact 906  
|||||  
Db 841 TGATTTGTGACTATGATTTATTTATTTCTACTATGTTTAAATGTTTATATGATAGTTTA 900  
QY 907 aagtaataataatgactagtagtagtagtagtagtagtagtagtagtagtagtagtag 966  
|||||  
Db 901 AAGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 960  
QY 967 attatgtgtgactaatcttataatcttcttcttcttcttcttcttcttcttcttct 1026  
|||||  
Db 961 ATTATGCTGACTTAATTTTCTATATTTTCTATATTTTCTATATTTTCTATATTTTCT 1020  
QY 1027 gtccgttaactagaattacagtggtggtccatgacactctggtgcttcttcttctatgcat 1086  
|||||  
Db 1021 GTCGTAACATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1080  
QY 1087 ggggtctgtgcgaagaaagaaagaaagaaagaaagaaagaaagaaagaaagaaagaa 1146  
|||||  
Db 1081 GGGGCTTGTGCGCAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 1140  
QY 1147 ggaatcaacaacaacaactcaaatagtcactggtgactgactgactgactgactgactgact 1206  
|||||  
Db 1141 GCAATACACACACCACTCAATTTATGTCAGTGGCTGATCAAGATCGCGCTGCATGATAT 1200  
QY 1207 gtctaaatgcatgcaagaagacacacgctgcttaacatgacatttaatggtccacacatct 1266  
|||||  
Db 1201 GTCTAAATGCGCATGCAAGACACGCTGTTAATCATGACACTTAAATGCTCACCACCATCT 1260  
QY 1267 caaccacacacacacacacatgctcttcttcttcttcttcttcttcttcttcttctt 1326  
|||||  
Db 1261 CAACCCACACACAAACACATGCTTTTCTTCTATCATCATCACACACACCTGTATATATAT 1320  
QY 1327 tcatctctctccgacactcaattcttctacttcaacaacagcttaaacctgataatgctg 1386  
|||||  
Db 1321 TCATTCCTTCGCGCACCTCAATTTCTTCACTTCAACACACGTCACCTGCATATGCGTG 1380  
QY 1387 tcatcccaatgcccacaaatctcatgcatgcttccacacacacacacacacacacacac 1446  
|||||  
Db 1381 TCATCCCATGCGCCAAATCTCATGCTGTTCCACACACACCTCTCTTATATATATATACCTA 1440  
QY 1447 taaataacttaataataactcaactcttcttcaatcaatcaatcaatcaatcaatcaatca 1506  
|||||  
Db 1441 TAAATACCTCTAATATATCATCTCTTCTATCATCATCATCATCATCATCATCATCATCT 1500  
QY 1507 tactactaataataacccaacccaactcataatcaataactcaactcaactcaactcaact 1557  
|||||  
Db 1501 TACTACTAT 1551

RESULT 3  
AX088019 1173 bp DNA linear PAT 17-MAR-2001  
LOCUS AX088019 Sequence 1 from Patent WO0114571.  
DEFINITION AX088019  
ACCESSION AX088019 GI:13396947  
VERSION AX088019.1  
KEYWORDS  
SOURCE  
ORGANISM  
Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;

REFERENCE 1 (bases 1 to 1173)  
AUTHORS van Rooijen, G., Keon, R. G., Boothe, J. and Shen, Y.  
TITLE Commercial production of chymosin in plants  
JOURNAL Patent: WO 0114571-A 1 01-MAR-2001;  
Sembiosys Genetics Inc. (CA)  
location/Qualifiers  
source 1. 1173  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
1. 1173  
CDS  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAC34680.1"  
/db\_xref="GI:13396948"  
/translation="MNP LKSPFPYAFICFGQYFVAVTHAAETRIPIYKGS LKALK  
EHGLDFLORQOIGISSKYSGRFVNSVPLNTLDSQYFGKITLGPIDPEFVLEDT  
GSSDFWVPSLYCKSNMCKNHRDPRKSRSTFONLGPLSLHYGDSMOGILGDTTVV  
SNIVDIOQTVGLSTOEPDQVETFAEPDGIIGMAYPSLASESLPDPNMNNRHLVAOD  
LFSYMDRNQESMLTLGAIDPSYTSLSHWPVTVQOYQOFVDSVTLGSVVACG  
GCAIIDTGTSLKVPSSDILNIDQATGATONOGDFRIDCDNLSYMPVPEFNGKM  
YPLRPSATYSQDQGFCTSGFQSENHRSKMLIGDYFIREYISVFDRAANNVLGAKAT"  
BASE COUNT 299 a 308 c 262 g 304 t  
ORIGIN  
Query Match 29.6%; Score 1173; DB 6; Length 1173;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1554 atgaacttccttaagctcttcccttcttcttcttcttcttcttcttcttcttcttcttctt 1613  
|||||  
Db 1 ATGAACCTCCTTAAGTCTTCCCTTTCTACACCTTCCCTTTGTTGCGTCAATFACTGCGTT 60  
QY 1614 gctgttactcaagctgctgagatcaacccgaattctctctcaagaaglaagcttctcgt 1673  
|||||  
Db 61 GCTGTACTCACGCTGCTGAGATCACCGCATTCCTCTCAAAAGTAACTCTCTCCGT 120  
QY 1674 aaggcgtcgaagaaacatgagactcttgaagaactcttgcgaagaacacagtaagcctc 1733  
|||||  
Db 121 AAGGCGCTGAAGAGACATGACCTTCTAGACACTTCTTGCAAGAACACACATATG6CATC 180  
QY 1734 agcagcaagtaactccgcttctgctgagatcttctgaagctgctgacattcaacacatcttcat 1793  
|||||  
Db 181 AGCAGCAAGTACTCCGCTTGGTGAAGTCTGTACGCGTCCACATACCAACTACTACTGAT 240  
QY 1794 agtcaatacttgggaagatctactctcgaaaccccgctcaagaagttcaacgcttctctt 1853  
|||||  
Db 241 AGTCAATCTTGGGAGAGATCTACTCTCGAACCCCGCTCAAGAGTTTACCCTCTCTTT 300  
QY 1854 gatactgttctctgacttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 1913  
|||||  
Db 301 GATACGTGCTCTCTGACTTCTGGGTTCTCTCTATCTACTGCAAGAGCAATGCTGCAAG 360  
QY 1914 aaccacaaagaatctcgatcgcgaagaagtcgctcacccttccgaactttaggaanaacccctt 1973  
|||||  
Db 361 AACCAACCAAGATTCGATCCGAGAAAGTCGTCACACTTCCAGAACTTAGGCAAAACCCCTTG 420  
QY 1974 tctatacaactaggttacaagtagcatgaagaacacttaggctatgataccgttactgtc 2033  
|||||  
Db 421 TCTATACACTACGGGTACAGGTAGCATGCAGGAATCTTAGGCTATGATACCGTCACTGTC 480  
QY 2034 tccacaacttggagacattcaacacagtagagactttagcaccacaaagacagtgatgtc 2093  
|||||  
Db 481 TCCAACATTTGGGACATTTCAACACACAGCTAGGACTTACACCCCAAGAACAGGATGATGTC 540  
QY 2094 ttcaacctatgcaaatctcgatgcatccttggtagttagcatcaacatcgctcgtcgtcagag 2153  
|||||  
Db 541 TTCACTATGAGAGATTCGATGCGATCCTTGTAATGACATACCATGCTGCGGTGACAG 600  
QY 2154 tactcgataactgtgttgaagaacatgataagacacacacacacacacacacacacacacac 2213  
|||||  
Db 601 TACTCGATACCTGTGTTGACACATGATGAACCGACACCTAGTAGCTCAAGACTTGTTC 660

QY 2214 tcggtttacatgagcaaggaaatgcccaggagagacatgctcaacgcttgagcatatgatcca 2273  
|||||  
Db 661 TCGGTTTACATGAGCAAGGAATGCGCAGAGAGCATGCTCACGCTTGAGGATATTGATCCA 720  
QY 2274 tctactacacagagatctctcaactgggttccagtcactgtgcagcaagtactggaattc 2333  
|||||  
Db 721 TCCACTATACAGAGATCTCTTCACTGGGTTCCAGTCACTGTGACACTACTGTGGCAATTC 780  
QY 2334 actgtgacagatgtcaccaataagcgtgtggtgtgtgtgcatgtgaagggtgagatgtcaact 2393  
|||||  
Db 781 ACTGTGACAGTGTACCATATAGCGGTGTGTTGTGCATGTGAAGGTGATGTCAAGCT 840  
QY 2394 atcttggatccggttagctgccaagcttggttcgagaccagcagacatctcaaatcag 2453  
|||||  
Db 841 ATCTGTGATACCGGTAGCTCCAGAGCTGTCGAGCTACAGAGACATTCCTCAACATTCAG 900  
QY 2454 caagcatatggagccacacagaaaccagtagcgtgagtttgacatagattgcgacaactc 2513  
|||||  
Db 901 CAAGCTATTGGAGGCCACACAGAACCCAGTACGTTGACATFAGATTGGGACAACTTT 960  
QY 2514 agctacagctacagatgtgtcttgagatcaacgagcaagatgtaccacatgacccctcc 2573  
|||||  
Db 961 AGCTACATGCTTACAGTTGTCTTTGAGATCAACGGCAGATGTACCCACTGACCCCTTC 1020  
QY 2574 gctctacacagccagatcaagggttctgcacacagtgatccagagtgagaaacattcc 2633  
|||||  
Db 1021 GCTATACACAGCAAGATCAAGGGTTCGACACAGTGATTCAGAGAGAACCATTC 1080  
QY 2634 cagaaatggatctctggagatgtgtcattcgtgtgagctacacagcttcttgacagggcc 2693  
|||||  
Db 1081 CAGAAATGGATCTTGGGAGATGTGTTCATTCGTGAGTACTACAGGCTTTGACAGGGCC 1140  
QY 2694 aacaacctcgttgagctagctaaagcaatctga 2726  
|||||  
Db 1141 AACAACTCTGTTGGGCTAGCTAAAGCAATCTGA 1173

RESULT 4  
AX343913 1415 bp DNA linear PAT 01-FEB-2002  
- LOCUS  
DEFINITION Sequence 5 from Patent WO0200899.  
ACCESSION AX343913  
VERSION AX343913.1 GI:18491959  
KEYWORDS  
SOURCE Phaseolus vulgaris.  
ORGANISM Phaseolus vulgaris.  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;  
Phaseolus.  
REFERENCE 1 (sites)  
AUTHORS Angenon, G., de Jaeger, G., Goossens, A. and Depicker, A.  
TITLE Heterologous gene expression in plants  
JOURNAL Patent: WO 0200899-A 5 03-JAN-2002;  
FEATURES location/Qualifiers  
source 1. 1415  
/organism="Phaseolus vulgaris"  
/db\_xref="taxon:3885"  
BASE COUNT 469 a 259 c 172 g 515 t  
ORIGIN

Query Match 26.5% Score 1047; DB 6; Length 1415;  
Best Local Similarity 99.6% Pred. No. 0;  
Matches 1367; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY 8 aatcatgttactcccaatcatcatatagtgaaaatttgctctctcgcggtgtttc 67  
|||||  
Db 44 AATCATGTACTCCCATATCATATATAGTGAAGTTTGGCTCTCTGCGCGTGTTT 103  
QY 68 ttacctctattaaagggtttcccaactaaaattctgtatcatctcacttacttg 127

Db 104 TTACCTCATTTAAAGGGGTTTCCACCTAAATAATTCGATATCTTCATCTTACTTG 163  
QY 128 ttacttattcttcataatcttggttgaataataacgcttccgcacagatactc 187  
|||||  
Db 164 TTACTTTAATTTCTCATATCTTGTGTGAATTTATTCAGGCTTCGACACAGATATTCCT 223  
QY 188 acaattctattatttgtttaaacatttccaacccgatalaaatttaltgaagtcgcgtcta 247  
|||||  
Db 224 ACAATTTATATTGTTTAAACATTTTCAACCCGATTAATTTATGAAAGTCCGCTTA 283  
QY 248 tcttaatgtgtctacaatttcaatltgaataataataatttacttaattttagcgttg 307  
|||||  
Db 284 TCTTTAATGTAGTCAATTTTCAATTTGAATTAATTAATTTACTTTATTTTACCGTTG 343  
QY 308 gtgagaagcataaagatttattcttattctcttcataataatgtttaataacata 367  
|||||  
Db 344 GTAGAAACCATATATATTATTTATTTCTTCTTCATTTAAATGTTTAATATACATATTA 403  
QY 368 acaaatcttctacactlaagaagatltcccatlttataatttlaaaaaatattatca 427  
|||||  
Db 404 AACAAATTTCTTACTTAAAGAAATTCACATTTATATTTTAAATATATTATATCA 463  
QY 428 ataatttcaaccagtaaatcctaataatagttgttccaagaatataaaatttaa 487  
|||||  
Db 464 ATATTTTCAACACAGTAAATCTCATTAATTAATTAATTTTCAAAAAGTAATAATTTTAA 523  
QY 488 ctccaaatttttttattcgcagctgacttlaagcaaacccagtgacaaacttagcaatt 547  
|||||  
Db 524 CTCCTAATTTTATTTATTCAGTGTATTAAGCAACCCAGTGACACACTACCCATT 583  
QY 548 ttcttcttgaataaaaaatccaatltacattgtatttttltatacaatgaataattc 607  
|||||  
Db 584 TTTTCTTTTGATTAATAAATCCAAATFATTCATTTGATTTTNTATACAAAGAAATTTTC 643  
QY 608 accaaacaatcattgtgtgtattcttgaagcaagtcagtatgcataatcattc 667  
|||||  
Db 644 ACCAAACATATTTGTGTATTTCTGAAGCAAGCATATGCAAAATTCATTAATTTTC 703  
QY 668 ccatttgcacactcaggaagtaacttgaagatctgcgttcttaactgcgagacacatcttcaa 727  
|||||  
Db 704 CCATTGACACTACGGAAGTATCTGAAGATCTGCTTTTACATGCGACACACTCTTTTAA 763  
QY 728 agtaatttataatagttactatataccaagattcatalataccaataactcaattact 787  
|||||  
Db 764 AGTATTTTAAATATAGTTACTATATTCACATATTCATATATCAAAATCTCAATTTACT 823  
QY 788 tctaaaaaattaatagatataataataatattacttctttaaatttgaattgaattgtc 847  
|||||  
Db 824 TCTAAAAATTAATATAGATTAATTAATTAATTAATTTTAAATTTTAAATTTTAAATTTGTT 883  
QY 848 gaattgtgactattgtatttattctactatgtttaaattgttltatagatgttaa 907  
|||||  
Db 884 GAATTTGTGACTATTTGATTTATTTATTTACTATGTGTTTAAATTTGTTTATAGTATTAA 943  
QY 908 agtaaatlaagtaatgtagtagagtgtagtagtglttaaccctaaaccaataactlaaca 967  
|||||  
Db 944 AGTAATATATAGTATGATAGTAGAGTGTAGAGTGTACCCCTAAACCATTAACCTATACA 1003  
QY 968 ttatgtgtgactaaatttccatataattctctaattgtccttcttacttcttctgtatgtaag 1027  
|||||  
Db 1004 TTTATGTGTGACTAATTTTCAATATATTTCTTATTTACCTTTTACCTTTTCTGTATGTAG 1063  
QY 1028 tccgttaactagaatagtagtggtgtagcactgtagctgtgtgttggatcagatg 1087  
|||||  
Db 1064 TCCGTATCTAGATTTACTGTGGGTGTCATGACACTCTGTGTGTTTGGTTCAATGCTAG 1123  
QY 1088 ggt-cttgcgaagaagaaagacaaagaaacaaagaaagaaacaaagagagagacaaac 1146  
|||||  
Db 1124 GATGCTTCGCAAGAAAGAACAAAGAACAAAGAAAGAAAGAAAGAACAAAGAGAGACAAAC 1183  
QY 1147 gcaatcacacaaccaactcaaattagtcaactggtcgtatcaagatcgcggttcaatgtat 1206  
|||||

```
Db 1184 GCATTCACACAAACCACTCAATTAAGTACGTGGCTGATCATGACCGCGTCCATGTAT 1243
Qy 1207 gcttaaatgcatgcaagaacaacagtgcttaacatgcatgcaacttaaaatggctacccatct 1266
Db 1244 GCTTAATATCCATGCAAAACACAGCTGCTTAACATGCACTTAAATGGCTCACCCATCT 1303
Qy 1267 caaccacacacaaacacacatgacctttcttctcatcatcacacacacacacgtatataat 1326
Db 1304 CAACCCACACACAAACACACTTGGCTTTTCTTCATCATCACACACACACACCTGTATATAT 1363
Qy 1327 tcatctctccgcgaacctcaattctctcatctcaacacacgtcaacctgca 1378
Db 1364 TCATCTCTTCCGCGACCTCAATTTCTTCACTTCAACACACGTCACACTGCA 1415

RESULT 5
AX252300 AX252300 2970 bp DNA linear PAT 05-OCT-2001
LOCUS AX252300
DEFINITION Sequence 2 from Patent WO0168887.
ACCESSION AX252300
VERSION AX252300.1 GI:15985641
KEYWORDS
SOURCE
ORGANISM synthetic construct.
REFERENCE
AUTHORS 1 (bases 1 to 2970)
TITLE Jung, R. and Kinney, A.J.
JOURNAL Hypoallergenic transgenic soybeans
Patent: WO 0168887-A 2 20-SEP-2001;
E. I. du Pont de Nemours and Company (US) ; PIONEER HI-BRED
INTERNATIONAL, INC. (US)
FEATURES
source 1..2970
location/Qualifiers
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Chimeric construct"
BASE COUNT 1018 a 547 c 527 g 878 t
ORIGIN
Query Match 22.4%; Score 887; DB 6; Length 2970;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 887; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2735 aagtgtaactaaatgcgtgaggtgtgtaaggtcattgtagagagcgtgtaaatgtatc 2794
Db 1796 AAGTATGAACCTAAATGACATGTAGGTGTAAGAGCTCATGAGAGCATGGAATATGTATTC 1855
Qy 2795 cgacacgttaacagtataaataactagctcaccatctcctctctatgataaacaag 2854
Db 1856 CGACCATGTAAACAGTATATATACTAGCTCCTCATCTCCTTCTCTATGATAAACAAG 1915
Qy 2855 atgtatgatataataaacctctatctatgcaacctattgtctctatgataaattcctct 2914
Db 1916 AGTATATGATATATTAACACTATCTATGACACTTATGTTCTATGATAAATTTCTCTCT 1975
Qy 2915 tatcttaaaatcctcgaatcgtgagagcgttatgagaaatgcttcaaatgatacaaaaac 2974
Db 1976 TATTTTATTAATCATCTGAAATCGTACGCTTATGGAATGCTTCAATATGTAACAATAAC 2035
Qy 2975 aaatgttactaataagacttctaaacaattcctaacttagactatgtgaacgagacata 3034
Db 2036 AAATGTGATCTATAGACTTCTTAAACATTTCTAAGTATGAGATTTGTAACGACATATA 2095
Qy 3035 gtgttaagaagacataacaataatgagaagatgtgtctccatctatataatataata 3094
Db 2096 GTTTTAAGAGACATTAACATTAATATGAGAGAGAGTTGTCTCCATTTATATATATATA 2155
Qy 3095 ttaccacattatgtattatattagagatgttgaaggagacataacaattataaagaagaag 3154
Db 2156 TTACCACATTATGTATATATATGAGATGTTAAAGAGACATTAACATTTATTAAGAGAGAG 2215
Qy 3155 ttgtatcatctatataatataactaccattatataatataatataatccactatatt 3214
```

```
Db 2216 TTTGTATTCATTTATATATATATATACCATTTATATATATATATATACCATTTATTT 2275
Qy 3215 aatgtcttataaaggttgatcacaatgatatttctaataatgaatgtagatgatga 3274
Db 2276 AATGTCTTATATAGGTTTATATCATGATATTTCTAATATATTTAGTTGATATATATATGA 2335
Qy 3275 aggtatctattgaactctctactctgataaaggttgatcacttaagttggtct 3334
Db 2336 AGGGTACTATTTGAACTCTTACTCTGTATTAAGGTGGATCATCTTAAAGTGGCTCT 2395
Qy 3335 atttaatttatatgtctctctacagataaaaaaaatlaagatggttggttgataaataat 3394
Db 2396 ATTTAATTTTATATGCTTCTTACAGATTAATAAAAAAATATATGATGCTTGAATAAATAT 2455
Qy 3395 tgaaggaatttaaaataataataataataacataacataatataatgatatataatct 3454
Db 2456 TGAAGGATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2515
Qy 3455 tataataacatttatactataaaaaaagaataatgataaataatctatacaatgctta 3514
Db 2516 TATATATTAACATTTATATGTAATAAAAAAGTAATATTTGCTATTAATCTATACATTCGTTTA 2575
Qy 3515 gccttgctgacgactcctaattatttaacgagagtaaacatatgtgacttttgta 3574
Db 2576 GCCTTGCTGAGACACTCTCATTTATTTAAACGAGACTTAACATATTTGACTTTTGCTTA 2635
Qy 3575 tttaacaattatatttaacactataagaatlttttttttalc 3621
Db 2636 TTAAACAATATATATTTAATTAACACTATATGAATTTTATTTTATATC 2682

RESULT 6
E00452
LOCUS E00452 1990 bp DNA linear PAT 29-SEP-1997
DEFINITION DNA sequence coding for faseorin.
ACCESSION E00452
VERSION E00452.1 GI:2168735
KEYWORDS E00452.1 GI:2168735
SOURCE JP 1985210988-A/2.
ORGANISM Escherichia coli.
Escherichia coli.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (bases 1 to 1990)
AUTHORS Jiyon,D.K., Yelmoshi,S.H., Jierli,E.S., Denisu,D.S. and
NORMOTO,M.
TITLE DEVELOPMENT OF PLANT STRUCTURAL GENE
JOURNAL Patent: JP 1985210988-A 2 23-OCT-1985;
AGURIJENETEKUSU RES ASSOC LTD
COMMENT OS Escherichia coli
PN JP 1985210988-A/2
PD 23-OCT-1985
PE 16-APR-1984 JP 1984077452
PR 15-APR-1983 US 83 485614
PI JIYON DEJI KENPU, TEIMOSHII SHII HOORU, JIERII ERU SURAITOMU,
PI DENISU DABURILYU SATSUMON, NORMOTO MURAI
PC C12N15/00,A01H1/00,C12N1/20,C12N5/00,(C12N1/20,C12R1:01), PC
(C12N5/00,
PC C12R1:91);
CC strandedness: Double;
CC topology: Linear;
CC hypothetical: No;
CC anti-sense: No;
CC "source": library-plasmid PKS4-KB;
FH key Location/Qualifiers
FT 5/UTR 1..177
FT CDS
FT join(178..490,563..753,842..922,1047..1277,1406..1664, FT
FT 1768..1958) /product="faseorin"
FT promoter 101..177
```

FT	exon	178. .490			
FT	intron	491. .562			
FT		/note='IVS1,72bp'			
FT	exon	563. .753			
FT	intron	754. .841			
FT		/note='IVS2,88bp'			
FT	exon	842. .922			
FT	intron	923. .1046			
FT		/note='IVS3,124bp'			
FT	exon	1047. .1277			
FT	intron	1278. .1405			
FT		/note='IVS4,128bp'			
FT	exon	1406. .1664			
FT	intron	1665. .1767			
FT		/note='IVS5,103bp'			
FT	exon	1768. .1958			
FT		1958. .<2090.			
FT	3'UTR				
FEATURES					
source					
Location/Qualifiers					
1. .1990					
/organism="Escherichia coli"					
/db_xref="taxon:562"					
BASE COUNT	614 a	416 c	372 g	588 t	
ORIGIN					

Query Match	4.6%	Score 181	DB 6	Length 1890
Best Local Similarity	100.0%	Pred. No. 9	9e-77	
Match 181	Conservative	0	Mismatches	0
			Indels	0
			Gaps	0

QY	1377	catatgggtatccatccatgcccacatctccatgacatgctccacacacactctctctat	1436
Db	1	CATATGGGTGTCATCCCATGCCCAATCTCCATGCATGTCTCCACACACTTCTCTTAT	60
QY	1437	ataatacctaataaactctaatalcactacactctttcatalcatalccatccagagt	1496
Db	61	ATATATACCTATAATACCTCTATATACACTCACTTCTTCATCATCATCCATCCAGAGT	120
QY	1497	actactactctactactataataatcccccaaccaactcatalatcatalactactctatg	1556
Db	121	ACTACTACTCTACTACTATATATACCCCAACCAACATCATATTCAATCACTACTACTATG	180
QY	1557	a 1557	
Db	181	A 181	

RESULT	7				
LOCUS	A06496	2090 bp	DNA	linear	PAT 10-NOV-1993
DEFINITION	Artificial sequence for phascolin.				
ACCESSION	A06496				
VERSION	A06496.1	GI:490411			
KEYWORDS					
SOURCE	synthetic construct.				
ORGANISM	artificial sequence.				
REFERENCE	1 (bases 1 to 2090).				
AUTHORS	Kemp,J.D., Hall,T.C., Slightom,J.L., Sutton,D. and Mural,N.				
TITLE	Plant structural gene expression				
JOURNAL	Patent: EP 0126546-A 24 28-NOV-1984;				
FEATURES	LUBRIZOL GENETICS INC				
Source	Location/Qualifiers				
	1..2090				
	/organism="synthetic construct"				
	/db_xref="taxon:32630"				
CDS	join(178..450,563..753,842..922,1047..1277,1406..1664				

```

/codon_start=1
/transl_table=11
/product="phaseolin"
/protein_id="CAA00562.1"
/db_xref="GI:490412"
translation="MMRRVPLLGLITLPLASLSAFATSLREEESQDNPFYNSDN

```

BASE COUNT	645 a	423 c	397 g	625 t
ORIGIN				
exon				<p> SNMTLTKKNVCGHTRVLORFPOOSKRLLONLEDYRLVFRSRSKPTLLLPQOADAELLVAV  RSSATLILVKKPRDDRREYFPLTSDNPFSDHOKIPAGTIFYLVNPDKEEDRIITOLAM  PVNNPQIHEFFFLSTEAQOOSTYLDQFSKHILEASFNSEFEEINLYLPEEGEOOGQIVYN  IDSEQIKELSKHAKSSRSKLSKODNTIGFGLIBRTNSLNLVLSIEMEGALF  VPRYSKAIVLIVNDEGAHVELVGPKNGETLEYEYSRAELSKDDEVLPAAIPVAI  KATSNVFTGFGGINANNRNRLLAGKTDNYTSSIGRALDGRDYGLTFSGSGDEVMKL  INFOSGSEYFVDANHHQEOEQKGRGAVV* </p>
intron				<p> &lt;178. .490  /number=1  491. .562 </p>
exon				<p> /number=1  563. .753 </p>
intron				<p> /number=2  754. .841 </p>
exon				<p> /number=2  842. .922 </p>
intron				<p> /number=3  923. .1046 </p>
exon				<p> /number=3  1047. .1277 </p>
intron				<p> /number=4  1278. .1405 </p>
exon				<p> /number=4  1406. .1664 </p>
intron				<p> /number=5  1665. .1767 </p>
exon				<p> /number=5  1768. &gt;1958 </p>

Query Match	4.6%	Score 181	DB 6	Length 2090
Best Local Similarity	100.0%	Pred. No.	9	8e-77
Matches 181	Conservative 0	Mismatches 0	Indels 0	Gaps 0

Qy	1377	catatggctgcatacccatcccaatctccatgagctgttccaaacaccttctccat	1436
Db	1	CATATGGCTGCAATCCCATCCCAATCTCCATGCATGTCCACCACTTCTCTTAT	60
Qy	1437	ataataacctataataacctataataataactactcttcttcaatcatalcatccagagt	1496
Db	61	ATAATACTTAATAATACCTCTTAATAATACACTCACTTCTTCATCATCATCATCCATCCAGAGT	120
Qy	1497	actactactactactactataataatacccaaccacatcatatcactactactatg	1556
Db	121	ACTACTACTACTACTACTATAATACCCCAACCAATCATATTCATCAATACTACTCTACTATG	180
Qy	1557	a	1557
Db	181	A	181

RESULT	8			
PHASE				
LOCUS		4764 bp	DNA	linear
DEFINITION		Phaseolus vulgaris gene for alpha-phaseolin.		
ACCESSION		X52626		
VERSION		X52626.1		
KEYWORDS		GI:20972		
SOURCE		alpha-phaseolin; glycoprotein; phaseolin; seed storage protein.		
ORGANISM		Phaseolus vulgaris.		
		Phaseolus vulgaris		
		Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
		Spermatophyta; Magnoliophyta; eudicotyledons: core eudicots;		
		Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae.		
		Phaseolus.		
		1 (bases 1 to 4764)		
REFERENCE		Anthony, J. L.		
AUTHORS		Direct Submission		
TITLE		Submitted (17-Apr-1990) Antony J. L., Dept. of Biology, Texas A & M		
JOURNAL		University, College Station, TX 77843-3258, USA		
REFERENCE		2 (bases 1987 to 4764)		

AUTHORS Anthony,J.L., Vonder Haar,R.A. and Hall,T.C.  
TITLE Nucleotide sequence of an alpha-phaseolin gene from Phaseolus vulgaris  
JOURNAL Nucleic Acids Res. 18 (11), 3396 (1990)  
MEDLINE 90287720  
FEATURES  
source Location/Qualifiers  
1..4764  
/organism="Phaseolus vulgaris"  
/strain="Sanilac"  
/db\_xref="taxon:3885"  
/clone="9c"  
/tissue\_type="leaf"  
2196..2207  
2232..2246  
2275..2664  
CAAT\_signal  
TATA\_signal  
exon  
CDS  
join(2352..2664,2737..2927,3016..3096,3221..3451,  
3580..3838,3942..4159)  
/product="alpha-phaseolin"  
/codon\_start=1  
/protein\_id="CA36853.1"  
/db\_xref="GI:295832"  
/db\_xref="SPTREMBL:Q41115"  
/translation="MMARVPLLLGLIFLASLSASFATSLREEESODNPFYNSDN  
SWNTLFKNQGHIRYLOREFDOOSKRLQMLEDYRLVEFRSKPETLLPQADAEILLVY  
RSGSAILVLVLPDDBREYFLTSDNPISDSDOKIPAGTIFLYVNDPDEDRITILAM  
PVNPDQIHFEFLSTEAQOSTLOERSKIIILASFSKREELIRVLPFEEGQDEGYIVN  
IDSEQIELSKHAKSSSRKSLSKODNTGNFNGINTEDNSLVNLSSIMKEGALF  
VPHYSKAVILVLVNEGEAHVELVGPKNKETLLEYSYRAELSKDQDVITPAAYPAI  
KATSNVNETGFGINANNNNRNLLAGKTDNVLSISIRALDGDVGLGTFSSGSGEYMKL  
INKSGSYFVDHGHQEOQKSHQEOQKGRKAFY"  
3552..2417  
2665..2736  
/number=1  
2737..2927  
/number=2  
2928..3015  
/number=2  
3016..3096  
/number=3  
3097..3220  
/number=3  
3221..3451  
/number=4  
3452..3579  
/number=4  
3580..3838  
/number=5  
3839..3941  
/number=5  
3942..>4159  
/number=6  
4084..4137  
/note="repeat region"  
BASE COUNT 1648 a 897 c 782 g 1437 t  
ORIGIN  
Query Match 4.3%; Score 169; DB 8; Length 4764;  
Best Local Similarity 99.3%; Pred. No. 6.2e-71;  
Matches 269; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1287 tgcctttctcatcatcacacaacacctgatatattatctcttcgcacacc 1346  
|||||  
DB 2085 tgcctttttcttcatcatcacacacacacctgatatattatctcttcgcacacc 2144  
QY 1347 aattcttcaacttcaacaacgtaacacctgcatgctgcatccatgccaatctc 1406  
|||||  
DB 2145 AATTCTTCACGTCAACACAGCTGAACCTGCATATGCCGTATCCCAATGC 2204  
QY 1407 catgcatgttccaacacctctctctataataactataaactctaatatcact 1466  
|||||  
DB 2205 CATGCATGTCCAACACACTCTCTCTATATATAATACCTATAAATACCCCTAATATCACT 2264

QY 1467 cacttcttcacatcatccatccacagagtactactactactactataataaccacac 1526  
|||||  
DB 2265 CACTTCTTTCATCATCATCCATCCACAGAGTACTACTACTACTACTATATAATACCCAAC 2324  
QY 1527 ccaactcatatcaataactactactactatga 1557  
|||||  
DB 2325 CCAACTCATATTCATCATACTACTACTACTATGA 2355  
RESULT 9  
LOCUS A11822 1475 bp mRNA linear PAT 12-NOV-1993  
DEFINITION Artificial mRNA for phaseoline.  
ACCESSION A11822  
VERSION A11822.1 GI:491197  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
1 (bases 1 to 1475)  
AUTHORS Kemp,J.D., Hall,T.C., Slightom,J.L., Sutton,D. and Murai,N.  
TITLE Patent: EP 0126546-A 33 28-NOV-1984;  
JOURNAL LUBRIZOL GENETICS INC  
FEATURES  
source Location/Qualifiers  
1..1475  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
78..1343  
/transl\_table=1  
/codon\_start=1  
/product="phaseoline"  
/protein\_id="CA00988.1"  
/db\_xref="GI:491198"  
/translation="MMARVPLLLGLIFLASLSASFATSLREEESODNPFYNSDN  
SWNTLFKNQGHIRYLOREFDOOSKRLQMLEDYRLVEFRSKPETLLPQADAEILLVY  
RSGSAILVLVLPDDBREYFLTSDNPISDSDOKIPAGTIFLYVNDPDEDRITILAM  
PVNPDQIHFEFLSTEAQOSTLOERSKIIILASFSKREELIRVLPFEEGQDEGYIVN  
IDSEQIELSKHAKSSSRKSLSKODNTGNFNGINTEDNSLVNLSSIMKEGALF  
VPHYSKAVILVLVNEGEAHVELVGPKNKETLLEYSYRAELSKDQDVITPAAYPAI  
KATSNVNETGFGINANNNNRNLLAGKTDNVLSISIRALDGDVGLGTFSSGSGEYMKL  
INKSGSYFVDHGHQEOQKSHQEOQKGRKAFY"  
BASE COUNT 455 a 334 c 306 g 380 t  
ORIGIN  
Query Match 3.4%; Score 133; DB 6; Length 1475;  
Best Local Similarity 100.0%; Pred. No. 2.6e-53;  
Matches 133; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2732 aataagatgaactaaatgatgtagtgaagagctcatgagagcattgaatatgt 2791  
|||||  
DB 1343 AATAAGTATGACATAAATGCATGTAGGTGAAGCCTCATGAGAGCATGAATATGT 1402  
QY 2792 atccgacacatgtaaacagataataactggtccatcttacttcttataaaca 2851  
|||||  
DB 1403 ATCCGACCATGTGAACAGTATAATACCTGAGCTCATCTCTTATGAATAACAA 1462  
QY 2852 aggatgtatgat 2864  
|||||  
DB 1463 AGCATCTTATGAT 1475  
RESULT 10  
PVPHASBR  
LOCUS PVPHASBR 1475 bp mRNA linear PLN 21-MAR-1995  
DEFINITION Phaseolus vulgaris mRNA for beta-type phaseolin.  
ACCESSION X03004  
VERSION X03004.1 GI:21039  
KEYWORDS glycoprotein; phaseolin; seed storage protein; signal peptide;  
storage protein.  
SOURCE Phaseolus vulgaris.

ORGANISM Phaseolus vulgaris  
Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;  
Phaseolus.  
1 (bases 1 to 1475)  
AUTHORS Slightom,J.L., Drony,R.F., Klassy,R.C. and Hoffman,L.M.  
TITLE Nucleotide sequences from phaseolin cDNA clones: the major storage  
proteins from Phaseolus vulgaris are encoded by two unique gene  
families  
JOURNAL Nucleic Acids Res. 13 (18), 6483-6498 (1985)  
MEDLINE 86041851  
COMMENT The single base deletion in a variant clone can be due either to a  
cloning artefact or to a represented pseudogene.  
FEATURES  
source 1. 1475  
/organism="Phaseolus vulgaris"  
/db\_xref="taxon:3885"  
misc\_RNA 1  
/note="cap site"  
1. 1475  
/note="mRNA"  
4  
/note="cap site in variant clone"  
78. 1343  
/note="precursor"  
/codon\_start=1  
/protein\_id="CAA26789.1"  
/db\_xref="GI:21040"  
/db\_xref="SWISS-PROT:P02853"  
/translation="MKRRPVLILGLIFLASLSASRATSLREEESODNPFYNSDN  
SWNTLFKNQGHIVLQRFDOQSRKILNEDYRLVERRSKPFETILLPQADAEILLVY  
RSGAILVLVFPDREVEFLTSDNPIFSDBQKIPAGTIFLVNPDREDRIILQIAM  
VNNPQJHEFLSTEAQSYLQEFKSHLEASFNSEFEINRVLFEEGQGEIVYN  
IDSDQIKELSHAKSSRSKSLSKODNTGNFGLTERDQSLNVLNLISSIMEGALF  
VPHYSKRAIVLYNCEGAHYEIVGPKNGTLEISYRAISKDDVPIPAAYPAI  
KAISVNFETGCIYANNNNRNLAKGTDNVTSIGRALDGDVIGLITFGSGDEVML  
INKQSGSYFVDAAHHHQEQKGRKGAIFY"  
78. 146  
/note="putative"  
147. 1340  
/product="mature beta-phaseolin (aa 1-398)"  
misc\_feature 149. 150  
/note="pot. altern. signal peptide cleavage site"  
152. 153  
/note="pot. altern. signal peptide cleavage site"  
155. 156  
/note="pot. altern. signal peptide cleavage site"  
170  
/note="A is missing in variant clone causing frameshift  
and premature stop codon"  
831. 839  
/note="pot. glycosylation site"  
1098. 1106  
/note="pot. glycosylation site"  
1434. 1458  
/note="put. polyadenylation signal"  
1475  
/note="polyadenylation site"  
BASE COUNT 455 a 335 c 306 g 379 t  
ORIGIN

Query Match 3.4%; Score 133; DB 8; Length 1475;  
Best Local Similarity 100.0%; Pred. No. 2.6e-53;  
Matches 133; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2732 aataagatgaactaaatgcatgtgtaagagcgaatgagagcgtgaatatgt 2791  
|||||  
Db 1343 AATAAGTATGAACATAAATGCAATGTAGTAAAGACTCATGAGAGCAATATTG 1402  
|||||  
Qy 2792 atcgcacatgttaacagtataataactggtccatctcaactcttctatgataaaca 2851  
|||||

Db 1403 ATCCGACCATGTACAGATATATAACTGAGCTCCATCTCACTTCTTCATGATATAACAA 1462  
Qy 2852 agagatgtatgat 2864  
|||||  
Db 1463 AGGATGTTATGAT 1475

RESULT 11  
ZMZEINI 207 bp DNA linear PLN 30-MAR-1995  
LOCUS Zee mays/P. vulgaris chimeric beta-phaseolin/zein gene 5' end region.  
DEFINITION X06175  
ACCESSION X06175.1 GI:22546  
VERSION X06175.1 GI:22546  
KEYWORDS phaseolin; signal peptide; zein protein.  
SOURCE Zee mays.  
Zea mays.  
Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
clade; Panicoideae; Andropogoneae; Zea.  
1 (bases 1 to 207)  
AUTHORS Hoffman,L.M., Donaldson,D.D., Bookland,R., Rashka,K. and  
Herman,E.M.  
TITLE Synthesis and protein body deposition of maize 15-kd zein in  
transgenic tobacco seeds  
JOURNAL EMBO J. 6, 3213-3221 (1987)  
COMMENT This sequence is a chimera between Phaseoleus vulgaris sequences  
from J.L. Slightom, S.M. Sun and T.C. Hall (1983) 'Complete  
nucleotide sequence of a French bean storage protein gene:  
phaseolin', Proc. Natl. Acad. Sci USA 80:1897-1901, and Zea mays  
sequences from  
K. Pedersen, P. Argos, S.V.L. Naravana and B.A. Larkins (1986)  
'Sequence analysis and characterization of a maize gene encoding a  
high-sulfur zein protein of Mr 15,000',  
J. Biol. Chem. 261:6279-6284.  
location/Qualifiers  
source 1. 207  
/organism="Zea mays"  
/db\_xref="taxon:4577"  
misc\_feature 1. 95  
/note="phaseolin sequence"  
9. 14  
/note="TATA element"  
41  
/note="phaseolin transcription initiation site"  
77. 81  
/note="TATA element"  
96. >207  
/note="zein sequence"  
98  
/note="zein transcription initiation site"  
161. >207  
161. >207  
/codon\_start=1  
/product="chimeric beta-phaseolin/zein"  
/protein\_id="CAA29542.1"  
/db\_xref="GI:22547"  
/translation="MKMYIVLVCAISAA"  
BASE COUNT 61 a 60 c 33 g 53 t  
ORIGIN

Query Match 2.4%; Score 95; DB 8; Length 207;  
Best Local Similarity 100.0%; Pred. No. 1.1e-34;  
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1437 ataatccataataatccctaatatcactcaactcttcacatcatcaccagagt 1496  
|||||  
Db 1 ATATATCCTATTAATACCTCTAATATACACTCACTTCTTCATCATCATCCATCCAGAGT 60  
|||||  
Qy 1497 actactactctactactataataatcccaaccaac 1531  
|||||  
Db 61 ACTACTACTCTACTACTATAATATACCCCAACCCAAC 95  
|||||

RESULT 12  
PVPHASAR  
LOCUS PVPHASAR 1535 bp mRNA linear PLN 21-MAR-1995  
DEFINITION Phaseolus vulgaris mRNA for alpha-type phaseolin.  
ACCESSION X02980  
VERSION X02980.1 GI:21035  
KEYWORDS direct repeat; glycoprotein; phaseolin; seed storage protein;  
signal peptide; storage protein; tandem repeat.  
SOURCE Phaseolus vulgaris.  
ORGANISM Phaseolus vulgaris.  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;  
Phaseolus.  
REFERENCE 1 (bases 1 to 1535)  
AUTHORS Slightom,J.L., Drong,R.F., Klassy,R.C. and Hoffman,L.M.  
TITLE Nucleotide sequences from phaseolin cDNA clones: the major storage  
proteins from Phaseolus vulgaris are encoded by two unique gene  
families  
JOURNAL Nucleic Acids Res. 13 (18), 6483-6498 (1985)  
MEDLINE 86041851  
COMMENT The variability in transcription initiation may be due to multiple  
and overlapping TATA-elements in the genomic sequence (see  
<PVPHASL2>). Seven slightly different alpha-phaseolin cDNA clones  
were detected. Whether the differences represent nucleotide  
substitutions in allelic genes or show divergences of non-allelic  
genes is unknown.  
FEATURES  
Source Location/Qualifiers  
1..1535 /organism="Phaseolus vulgaris"  
/db\_xref="taxon:3885"  
1 /note="cap site"  
misc\_RNA 10  
/note="cap site in variant clones"  
misc\_RNA 13  
/note="cap site in variant clone"  
CDS 87..1397  
/note="precursor"  
/codon\_start=1  
/protein\_id="CAA26718.1"  
/db\_xref="GI:21036"  
/db\_xref="SWISS-PROT:P07219"  
/translation="MMARVPLLLGLTFLASLSASPATSLREEESODNPFFNSN  
SMNTFKNQYGHIFVLORPDOQSKRLONLEDRYVERRSKRETLILQQDAELLIV  
RSGSAILVLYKPPDRREFFLTQGNFTSDNQKTPAGTITFLVNPDKEDLRIQLA  
MPVNPQIHFEFLSTEAQOQSYLOEFKHLLEAFNSKFEELNVLEEEQOEQGO  
EGVIVNDSQIELSKHAKSSRSKSHKODNTIGNFNGTERTDNLNLLISIEH  
KEGALFVPHYSKAIVILVYNEGFAHVELGPKNKELEPESYRAELSKDQVPIPA  
AYPVAKATSNVNTFGINANNNNRNLLAGKTNYVLSIGRALDGDVGLTFSGSG  
EYWKILNKSGSFYVDGHHHQDQDQKSHQDQKRGKAFV"  
sig\_peptide 87..155  
/note="putative"  
mat\_peptide 156..1394  
/product="put. mature alpha-phaseolin (aa 1-413)"  
misc\_feature 158..159  
/note="pot. altern. signal peptide cleavage site"  
misc\_feature 161..162  
/note="pot. altern. signal peptide cleavage site"  
misc\_feature 164..165  
/note="pot. altern. signal peptide cleavage site"  
variation 339  
/note="C is T in variant clones"  
variation 456..464  
/note="ACGCAAGC (Thr, Gln, Gly) is ACCAGC (Thr, Ser) in  
variant clones"  
variation 456..464  
/note="ACGCAAGC (Thr, Gln, Gly) is GCTAGC (Ala, Ser) or  
ACTAGC (Thr, Ser) in variant clones"  
variation 486..488  
/note="AAC (Asn) is CAC (His) in variant clones"  
variation 533

variation /note="C is T in variant clones"  
594..596  
/note="GAA (Glu) is GAC (Asp) in variant clones"  
repeat\_region 711..725  
/note="direct repeat 1"  
repeat\_region 726..740  
/note="direct repeat 1"  
misc\_feature 858..866  
/note="pot. glycosylation site"  
misc\_feature 1125..1133  
/note="pot. glycosylation site"  
repeat\_region 1321..1348  
/note="direct repeat 2"  
repeat\_region 1349..1375  
/note="direct repeat 2"  
misc\_feature 1509..1514  
/note="put. polyadenylation signal"  
misc\_feature 1530  
/note="polyadenylation sites in variant clones"  
BASE COUNT 480 a 344 c 323 g 388 t  
ORIGIN

Query Match 2.3%; Score 90; DB 8; Length 1535;  
Best Local Similarity 100.0%; Pred. No. 2,4e-32;  
Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1468 actcttcacatccatccatccagagtactactactactactataatacccccaacc 1527  
|||||  
DB 1 ACTCTTTCATCATCCATCCATCCAGAGTACTACTACTACTATAATACCCCAACC 60  
OY 1528 caactcatatcaatcactactactatga 1557  
|||||  
DB 61 CAACTCATTTCAATCACTACTACTACTATGA 90

RESULT 13  
LOCUS U01131 1454 bp mRNA linear PLN 16-MAY-1996  
DEFINITION Phaseolus vulgaris Sanilac clone 1-12 phaseolin (Phs) mRNA,  
complete cds.  
ACCESSION U01131  
VERSION U01131.1 GI:403593  
KEYWORDS  
SOURCE  
ORGANISM French bean.  
Phaseolus vulgaris  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;  
Phaseolus.  
REFERENCE 1 (bases 1 to 1454)  
AUTHORS Kaml,T.A. and Gepts,P.  
TITLE Phaseolin nucleotide sequence diversity in Phaseolus. I.  
Intraspecific diversity in Phaseolus vulgaris  
JOURNAL Genome 37, 751-757 (1994)  
MEDLINE 95095072  
REFERENCE 2 (bases 1 to 1454)  
AUTHORS Kaml,T.A.  
TITLE Direct Submission  
JOURNAL Submitted (30-AUG-1993) James A. Kaml, Agronomy and Range Science,  
University of California at Davis, Davis, CA 95616, USA  
FEATURES  
Source Location/Qualifiers  
1..1454  
/organism="Phaseolus vulgaris"  
/cultivar="Sanilac"  
/db\_xref="taxon:3885"  
/map="linkage group D7"  
/clone="Sanilac 1-12"  
/haplotype="S" type beta-phaseolin"  
/tissue\_type="cotyledon"  
/clone\_lib="Sanilac PUC19 cDNA"  
/dev\_stage="cotyledonary stage"  
1..1454

gene

```

5'UTR      /gene="Pns"
           1..53
           /gene="Pns"
           54..1319
           /gene="Pns"
           /function="seed storage protein"
           /codon_start=1
           /product="phaseolin"
           /protein_id="AA09534.1"
           /db_xref="GI:403594"
           /translation="MVARVPLMLGLFLASLSASFATSLREEESODNPEFNSDN
           SNNRLPKNOYGHTRVRORPDOOSKLONIEDVRLVFRSKPETLLIPQADALLVY
           KSGSALIVLKPDDRREYFFLSDNPFSDHOKTPAGTIFYLYNPDPKEDLRIOLAM
           PVNNPQIHDFLSTEAQSYLQEFKSHLEASFNSKFEETNVLFEESQDEGVIN
           IDSEQIKELSKHAKSSRSKLSKQDNTIGNFENGLNTERDNLNLNVLFEESQDEGVIN
           VPHYYSKATYIIVNNEGAEHVELGPKNGKTELEYEYRAELSKDVFVIPAAYPAI
           KATSNVPTFGTGINANNNNNLGAKTDNYSISGRALDKDVLGLTFSSGDEVNKL
           MKQSSYFVDHHDHOOEOKGKGAIFY"
sig_peptide 54..125
           /gene="Pns"
           123..1316
           /gene="Pns"
           /function="seed storage protein"
           /product="phaseolin"
           674..679
           /gene="Pns"
           /standard_name="T' type alpha-phaseolin"
           /label=15bp-repeat
           /phenotype="T' type phaseolin"
           /replace="gagagaggagacagaagagagagagagca"
           /product="phaseolin"
           807..815
           /gene="Pns"
           /function="putative glycosylation site"
           1074..1082
           /gene="Pns"
           /function="putative glycosylation site"
           1271..1298
           /gene="Pns"
           /standard_name="T' type alpha-phaseolin"
           /label=27bp-repeat
           /phenotype="T' type phaseolin"
           /replace="tcaccaacaggaacagaagaaggaaatcaccacaaggaacagaagaag
           gga"
           /product="phaseolin"
           1320..1454
           /gene="Pns"
           1429..1434
           /gene="Pns"
           3'UTR      polyA_signal
BASE COUNT 451 a 324 c 306 g 373 t
ORIGIN
Query Match      1.9%; Score 77; DB 8; Length 1454;
Best Local Similarity 100.0%; Pred. No. 5.2e-26;
Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 2769 tgtatccagcatgtcaacagataataactgagccatctcaactcttctatagaataa 2848
|||||
Db 1375 tctatccagcatgtcaacagataataactgagccatctcaactcttctatagaataa 1434
|||||
Oy 2849 caaagagatgtatgata 2865
|||||
Db 1435 CAAAGAGATGTATGATA 1451
|||||
RESULT 14
ZMZEIN2      ZMZEIN2      348 bp DNA linear PLN 06-AUG-1992
LOCUS        Maize chimeric zein/beta-phaseolin gene 3'end region.
DEFINITION   X06176
ACCESSION    X06176.1 GI:22548
VERSION      phaseolin; zein protein.
KEYWORDS

```

```

SOURCE      Zea mays.
ORGANISM    Zea mays.
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1 (bases 1 to 348)
AUTHORS     Hoffmann,L.M., Donaldson,D.D., Bookland,R., Rashka,K. and
            Herman,E.M.
TITLE        Synthesis and protein body deposition of maize 15-kd zein in
            transgenic tobacco seeds
JOURNAL     EMBO J. 6, 3213-3221 (1987)
FEATURES    Location/Qualifiers
            source          1..348
                           /organism="Zea mays"
                           /db_xref="taxon:4577"
                           <1..269
                           /note="zein sequence"
                           192..197
                           /note="polyA signal"
                           236..241
                           /note="polyA signal"
                           253..258
                           /note="polyA signal"
                           270..348
                           /note="phaseolin sequence"
                           298..303
                           /note="polyA signal"
                           338
                           /note="pot.polyA site"
                           340
                           /note="pot.polyA site"
polyA_site  93 a 93 c 62 g 100 t
BASE COUNT  93 a 93 c 62 g 100 t
ORIGIN
Query Match      1.9%; Score 75; DB 8; Length 348;
Best Local Similarity 100.0%; Pred. No. 5.9e-25;
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 2819 gagctcatctcaactcttctcatgaataaagaagatgtatgatataatacaccta 2878
|||||
Db 274 GAGCTCATCTCACCCTCTCTATGATTAACAAAGAGATGATATTAACACTCTA 333
|||||
Oy 2879 tctatgacacctat 2893
|||||
Db 334 TCTATGACACCTTAT 348
|||||
RESULT 15
LOCUS        U01121      1478 bp mRNA linear PLN 16-MAY-1996
DEFINITION   Phaseolus lunatus phaseolin (Pns) mRNA, complete cds.
ACCESSION    U01121
VERSION      U01121.1 GI:403581
KEYWORDS
SOURCE       Lima bean.
ORGANISM     Phaseolus lunatus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Phaseolus.
REFERENCE   1 (bases 1 to 1478)
AUTHORS     Kaml,U.A. and Gepts,P.
TITLE        Phaseolin nucleotide sequence diversity in Phaseolus. I.
            Intraspecific diversity in Phaseolus vulgaris
JOURNAL     JOURNAL MEDLINE
            Genome 37, 751-757 (1994)
95095072
REFERENCE   2 (bases 1 to 1478)
AUTHORS     Kaml,U.A.
TITLE        Direct Submission
JOURNAL     Submitted (30-AUG-1993) James A. Kaml, Agronomy and Range Science,
            University of California at Davis, Davis, CA 95616, USA
            Location/Qualifiers

```



```
source      1..1478
            /organism="Phaseolus lunatus"
            /strain="G25804"
            /db_xref="taxon:3884"
            /clone="G25-39"
            /tissue_type="cotyledon"
            /clone_lib="G25 pUC19 cDNA"
            /dev_stage="cotyledonary stage"
            1..1478
            /gene="Phs"
            1..60
            /gene="Phs"
            61..1347
            /gene="Phs"
            /function="seed storage protein"
            /codon_start=1
            /product="Phaseolin"
            /protein_id="AA09532.1"
            /db_xref="GI:403582"
            /translation="MRAVPLLLGLFLASLSPALISLREHNESQDNPFFSSDN
            SMOTLFKNQGHIVLOSFDQSERLQNLDPYRVEPMSKPEITLLPOADAPELIVY
            RSGSALLALVYKPGSTIYSLKQDPTLKIPACTITPFLINPQNNEDIRITKLMATVNNPQ
            IODFLSSTEAOQSYLIGFRKDIIDASFNSPIEETIRLLFAEGKQEGVIVNIGSDLI
            QELSRHAKSSSRKSLDHSNLDISNEMGMLTDIVNSLDVLTYYVEIKEGGLFVPHYS
            KAIYILVEEGVAKEVEIYGPKEKSELEETRADYSEGDFVIIPAAVPAIKAISNV
            NTFSEGINANNRYRILLTGKGPGKEDNIIISAGINPDVGLMFPGSGEDVQKLFNNQ
            NLSHFVNGSYHKNNQPOHQEQOKQKGRKAFVY"
            61..132
            /gene="Phs"
            133..1344
            /gene="Phs"
            /function="seed storage protein"
            /product="phaseolin"
            151..159
            /gene="Phs"
            /function="putative glycosylation site"
            796..804
            /gene="Phs"
            /function="putative glycosylation site"
            1063..1071
            /gene="Phs"
            /function="putative glycosylation site"
            1237..1245
            /gene="Phs"
            /function="putative glycosylation site"
            1255..1263
            /function="putative glycosylation site"
            /gene="Phs"
            /function="putative glycosylation site"
            1348..1478
            /gene="Phs"
            1457..1462
            /gene="Phs"
            BASE COUNT  441 a 340 c 311 g 386 t
            ORIGIN

Query Match      1.6%; Score 63; DB 8; Length 1478;
Best Local Similarity 100.0%; Pred. No. 3.4e-19;
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1495 gtactactactactataatacccccaaccacatatacatatactactacta 1554
      |||
Db 2  gTACTACTACTACTATAATACCCACCAACCATATTCATATCAATACACTACTACTA 61
QY 1555 tga 1557
      |||
Db 62  TGA 64
```

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 1, 2002, 01:57:49 ; Search time 791.3 Seconds  
(without alignments)  
8585.657 Million cell updates/sec

Title: US-09-643-755B-3

Perfect score: 3957

Sequence: 1 ctgcagatcatctgtact.....acctcaactaaggtacc 3957

Scoring table:

OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 1736436 seqs, 858457221 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

N.Geneseq\_032802.\*  
1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*  
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*  
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*  
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*  
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*  
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*  
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*  
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*  
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*  
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*  
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*  
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*  
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*  
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*  
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*  
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*  
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*  
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*  
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*  
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*  
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*  
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*  
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*  
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3957	100.0	3957	22	AA500570
2	1544	39.0	1558	20	AA10376
3	1439	36.4	1547	20	AA930961
4	1236	31.0	1244	20	AA10392
5	1173	29.6	1173	22	AA500569
6	887	22.4	2970	22	AA017528
7	49	1.2	1098	4	AA130063
8	49	1.2	1098	11	AA004683
9	49	1.2	1175	5	AA040295

10	49	1.2	1175	13	AA020949
11	49	1.2	1210	12	AA014051
12	49	1.2	1240	16	AA030006
13	49	1.2	1278	5	AA040055
14	49	1.2	1289	4	AA030022
15	49	1.2	1291	10	AA091157
16	49	1.2	1314	4	AA030049
17	49	1.2	1460	3	AA020043
18	49	1.2	1460	5	AA040180
19	49	1.2	2727	10	AA091188
20	49	1.2	2733	20	AA0206463
21	49	1.2	2982	10	AA091185
22	38	1.0	1290	4	AA030209
23	35	0.9	637	14	AA049459
24	34	0.9	1096	20	AA083966
25	33	0.8	179	5	AA040296
26	33	0.8	186	13	AA020950
27	33	0.8	819	20	AA083967
28	32	0.8	107	12	AA014777
29	29	0.7	1460	3	AA020043
30	29	0.7	1460	5	AA040180
31	25	0.6	33	21	AA062421
32	24	0.6	25	20	AA010389
33	23	0.6	1083	21	AA097159
34	23	0.6	1140	24	AA044658
35	23	0.6	1814	21	AA037875
36	23	0.6	44242	23	AB019930
37	22	0.6	262	22	AB07288
38	22	0.6	262	22	AA089933
39	22	0.6	3607	18	AA089930
40	22	0.6	5376	24	AB014150
41	22	0.6	6385	22	AA045352
42	22	0.6	14708	22	AA045513
43	22	0.6	14708	22	AA046773
44	22	0.6	14708	24	AB034244
45	22	0.6	15881	24	AB032260

#### ALIGNMENTS

RESULT 1  
AAS00570 standard; DNA; 3957 BP.

AC AAS00570;

DT 14-MAY-2001 (first entry)

DE Bovine phaseolin promoter pre-pro-chymosin-phaseolin terminator.

XX Chymosin; transcription regulator; terminator sequence; soybean; corn;  
pre-pro-chymosin; rape seed; sunflower; cotton; tobacco; alfalfa; wheat;  
barley; oats; sorghum; Arabidopsis thaliana; potato; flax; linseed; rice;  
safflower; oil palm; ground nut; Brazil nut; coconut; castor; coriander;  
squash; jojoba; ds; phaseolin; promoter; terminator; mutant; French bean.

XX Chimeric - Bos sp.

OS Chimeric - Phaseolus vulgaris.

XX Location/Qualifiers

FT Key

FT promoter

FT CDS

FT terminator

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

Prochymosin (prore  
Renin gene. Synt  
Chymosin open read  
Sequence of prochy  
Sequence of veal c  
Cloned sequence of  
cDNA sequence corr  
Pre-prorennin-A ge  
Sequence of recomb  
BamHI/Sall insert  
2.7 Kbp HindIII fr  
BamHI insert from  
Sequence of p1epro  
Prochymosin gene 5  
DNA encoding the f  
Sequence coding fo  
DNA coding for a r  
DNA encoding His-c  
Renin - casein co  
Pre-prorennin-A ge  
Sequence of recomb  
Phascolin gene pro  
PCR primer used to  
Arabidopsis thalia  
Human aspartyl pro  
Arabidopsis thalia  
Drosophila melanog  
Human pancreatic c  
Human digestive sy  
Cryptosporidium pa  
Human immune syste  
Chemically pretrea  
Tumour suppressor  
Human immune syste  
Human immune syste

PD	01-MAR-2001.
XX	
PF	23-AUG-2000; 2000OWO-CA00975.
XX	
PR	23-AUG-1999; 99US-0378696.
XX	
PA	(SEMBIOSIS GENETICS INC.
XX	
PI	Van Rooijen G, Keon RG, Boothe J, Shen Y;
DR	WPI; 2001-226621/23.
P-PSDB:	AAU00536.
XX	
PT	Producing chymosin in seeds of plants such as rice, flax, rape seed, by
PT	transforming plant cell with a nucleic acid encoding chymosin operably
PT	linked to transcription regulator and terminator sequences -
XX	
PS	Example 1; Fig 2; 56pp; English.
XX	
CC	The sequence represents a chimeric polynucleotide comprising a
CC	pre-pro-chymosin, a phaseolin promoter and a phaseolin terminator.
CC	Chymosin can be produced in a plant seed through introduction of a
CC	chimeric nucleic acid molecule, comprising a nucleic acid sequence
CC	encoding a chymosin polypeptide operatively linked to transcription
CC	regulator and terminator sequences, in particular a plant cell. The sequences are
CC	useful for producing plant seeds, in particular seeds of soybean, rape
CC	seed, sunflower, cotton, corn, tobacco, alfalfa, wheat, barley, oats,
CC	sorghum, Arabidopsis thaliana, potato, flax/lineded, safflower, oil palm,
CC	groundnut, Brazil nut, coconut, castor, coriander, squash, jojoba and
CC	rice.
XX	
SQ	Sequence 3957 BP, 1263 A; 790 C; 609 G; 1295 T; 0 other;
XX	
Query Match	100.0%; Score 3957; DB 22; Length 3957;
Best Local Similarity	100.0%; Pred. No. 0;
Matches 3957; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
OY	1 ctgcaggaatcatgttactccagcaglatcataagtgaagglttgctctcgccg 60
DB	1 ctgcaggaattcatgttactccagcaglatcataagtgaagglttgctctcgccg 60
OY	61 tggttttactcttatttaagggggtttccacctaanaatttcggtatactttcact 120
DB	61 tggttttactctcatttaagggggtttccaccctaaaatcttcggtatactttcact 120
OY	121 ttacttgttacttaatttctcatalaactcttggttgaattacaagcttcgcacaga 180
DB	121 ttacttgttacttaatttctcatalaactcttggttgaattacaagcttcgcacaga 180
OY	181 tatccctaaaatttatatttggtaaagaatttcaaacgcgataaattttatgaagtc 240
DB	181 tatccctaaaatttatatttggtaaagaatttcaaacgcgataaattttatgaagtc 240
OY	241 ccgtctatctttaaagtagtccaacatttcaatigaataataataatttacttaattt 300
DB	241 ccgtctatctttaaagtagtccaacatttcaatigaataataataatttacttaattt 300
OY	301 agcgttgtagaagaagcataaagattatcttactcttcttcataataaagtftaataba 360
DB	301 agcgttgtagaagaagcataaagattatcttactcttcttcataataaagtftaataba 360
OY	361 caataaacaatcttaccttaagaaggattcccatltaatttaataataatat 420
DB	361 caataaacaatcttaccttaagaaggattcccatltaatttaataataatat 420
OY	421 ttatcaaatatttticaaacgcgtaaatctcaataataaagtgttcaaaagtatbaa 480
DB	421 ttatcaaatatttticaaacgcgtaaatctcaataataaagtgttcaaaagtatbaa 480
OY	481 aatttaactccataattttttatctcgactgatcttaagaacaaccagtgacaact 540
DB	481 aatttaactccataattttttatctcgactgatcttaagaacaaccagtgacaact 540

QY	541	agcatcttttctcttggaataaaaaaacccaattacatctgatttttttatcaatga	600
Db	541	agccaatttttctcttggaataaaaaaacccaattacatctgatttttttatcaatga	600
QY	601	aaattcaccaaaaacatcttcttggttatcttcgaaagcaatgatactgataaatct	660
Db	601	aaattcaccaaaaacatcttcttggttatcttcgaaagcaatgatactgataaatct	660
QY	661	ataattccatttgcacactacggaagtacatgaagatcgtctttaaactgagacacat	720
Db	661	ataattccatttgcacactacggaagtacatgaagatcgtctttaaactgagacacat	720
QY	721	cttctaaagtaattttaaataatgatttaccatctatctaaagtttccatatatcaatccaa	780
Db	721	cttctaaagtaattttaaataatgatttaccatctatctaaagtttccatatatcaatccaa	780
QY	781	tattactcttaaaaaataattagataataatlaaataattacttttttaatttaagt	840
Db	781	tattactcttaaaaaataattagataataatlaaataattacttttttaatttaagt	840
QY	841	aattggtgaatttcttgacatttgatttatatttctactatggttaatttggttatgat	900
Db	841	aattggtgaatttcttgacatttgatttatatttctactatggttaatttggttatgat	900
QY	901	agtttaaagtaataataagaatgacgtatgagtggttagatgttaccataacataaac	960
Db	901	agtttaaagtaataataagaatgacgtatgagtggttagatgttaccataacataaac	960
QY	961	tataacattatggttggaactaatlttccatatatttcttatgtctttaacttccgtgt	1020
Db	961	tataacattatggttggaactaatlttccatatatttcttatgtctttaacttccgtgt	1020
QY	1021	atgtaagtcgctaactcgtgaattacagtttggttgccatggaactcgtgtcttttgctc	1080
Db	1021	atgtaagtcgctaactcgtgaattacagtttggttgccatggaactcgtgtcttttgctc	1080
QY	1081	atgcatagtgtcttcgcgaagaaaaaagacaagaacaagaaaaaagacaaaaacagagaga	1140
Db	1081	atgcatagtgtcttcgcgaagaaaaaagacaagaacaagaaaaaagacaaaaacagagaga	1140
QY	1141	caaaaacgaatcacacaacaacactcaaatgactgactggtgcataagaatcgccgctgc	1200
Db	1141	caaaaacgaatcacacaacaacactcaaatgactgactggtgcataagaatcgccgctgc	1200
QY	1201	atgtaatgcttaaatgacatgacaaagaagaacagtgcttaacatgacactttaaigtgcac	1260
Db	1201	atgtaatgcttaaatgacatgacaaagaagaacagtgcttaacatgacactttaaigtgcac	1260
QY	1261	coactccaacccacacacaacacacatctgtccttcttccatcatcacacaacaacactgt	1320
Db	1261	coactccaacccacacacaacacacatctgtccttcttccatcatcacacaacaacactgt	1320
QY	1321	ataatatcatctcttccgccacccctaatttcttcaacttcaacacacgltcaactgtcata	1380
Db	1321	ataatatcatctcttccgccacccctaatttcttcaacttcaacacacacgltcaactgtcata	1380
QY	1381	tgcgtgtcatcccatgcccgaactctccatgcatggtttccaacacacttctcttatataa	1440
Db	1381	tgcgtgtcatcccatgcccgaactctccatgcatggtttccaacacacttctcttatataa	1440
QY	1441	tacctataaatacctccaataataatcatcactcttcttccatcatcccatcgaagtata	1500
Db	1441	tacctataaatacctccaataataatcatcactcttcttccatcatcccatcgaagtata	1500
QY	1501	ctactctactactataataccaccaacccaactcataatccaataactactctatgaaact	1560
Db	1501	ctactctactactataataccaccaacccaactcataatccaataactactctatgaaact	1560
QY	1561	tccttaagtcttcccttcttaagcttcccttggtttgttggtcaaatcgtgtgtgtgta	1620
Db	1561	tccttaagtcttcccttcttaagcttcccttggtttgttggtcaaatcgtgtgtgtgta	1620

QY 1621 ctcaagctgctgagatcaaccgcatctctctatacaaaagtcaatgtcttcgtaagcgc 1680  
|||||  
Db 1621 ctcaagctgctgagatcaaccgcatctctctatacaaaagtcaatgtcttcgtaagcgc 1680  
QY 1661 tgaaggaaacatgagactctcttaagaactctctgcagaaacaacagtatgcatcagacga 1740  
|||||  
Db 1661 tgaaggaaacatgagactctcttaagaactctctgcagaaacaacagtatgcatcagacga 1740  
QY 1741 agtactccggctcggtgaagtctgctagcgtgcacacttacaacacttataatgaat 1800  
|||||  
Db 1741 agtactccggctcggtgaagtctgctagcgtgcacacttacaacacttataatgaat 1800  
QY 1801 acttgggaagaatctacactcgcgaaccccgccctcaagagtcaacgcttctcttgaatg 1860  
|||||  
Db 1801 acttgggaagaatctacactcgcgaaccccgccctcaagagtcaacgcttctcttgaatg 1860  
QY 1861 gtctcctctgaactcttggtgttcctctatactatctgaagaagcaatgctgcagaacaac 1920  
|||||  
Db 1861 gtctcctctgaactcttggtgttcctctatactatctgaagaagcaatgctgcagaacaac 1920  
QY 1921 aaagattcgatccgagaagaagtcgtccacacttccagaacttgggcaaacctgtctctaac 1980  
|||||  
Db 1921 aaagattcgatccgagaagaagtcgtccacacttccagaacttgggcaaacctgtctctaac 1980  
QY 1981 actacggtacagagtacgaatgcaatgcaagaatcttaagctatgataccgctacgtctccaca 2040  
|||||  
Db 1981 actacggtacagagtacgaatgcaatgcaagaatcttaagctatgataccgctacgtctccaca 2040  
QY 2041 ttgtggagacttcaacagacagttagacttagaccaccaagaaacaggtgtagtcttcaccc 2100  
|||||  
Db 2041 ttgtggagacttcaacagacagttagacttagaccaccaagaaacaggtgtagtcttcaccc 2100  
QY 2101 atgacagaaatctgacagatcgacactctgtatgcatataccactgcgtccgacagatctcga 2160  
|||||  
Db 2101 atgacagaaatctgacagatcgacactctgtatgcatataccactgcgtccgacagatctcga 2160  
QY 2161 taacctgtgttgacaacaatgatacgaacacactagtagctcaagaactgtgtctcgttt 2220  
|||||  
Db 2161 taacctgtgttgacaacaatgatacgaacacactagtagctcaagaactgtgtctcgttt 2220  
QY 2221 acatggagagaatggtgcagagaagacatgctcaacgcttggagctatgtatccactact 2280  
|||||  
Db 2221 acatggagagaatggtgcagagaagacatgctcaacgcttggagctatgtatccactact 2280  
QY 2281 acacaggaatcctctcaactggttccagctactgtgcagcagtaactggaattcactgtg 2340  
|||||  
Db 2281 acacaggaatcctctcaactggttccagctactgtgcagcagtaactggaattcactgtg 2340  
QY 2341 acagtgctcaacatcagcgtgtgtgtgttgcatgtgaaagtgtagtcaagctatctctg 2400  
|||||  
Db 2341 acagtgctcaacatcagcgtgtgtgtgttgcatgtgaaagtgtagtcaagctatctctg 2400  
QY 2401 ataccggtacgttccaagctgtgcagacttagcagcagacatctcaacatltcagaagaacta 2460  
|||||  
Db 2401 ataccggtacgttccaagctgtgcagacttagcagcagacatctcaacatltcagaagaacta 2460  
QY 2461 ttggagacacacagaaacagatgagtttgacatagatltgcagacaacttgaactga 2520  
|||||  
Db 2461 ttggagacacacagaaacagatgagtttgacatagatltgcagacaacttgaactga 2520  
QY 2521 tgcctacagttgtcttgaagatcaacgagcaagatgtagtcaacacacacacacacacacac 2580  
|||||  
Db 2521 tgcctacagttgtcttgaagatcaacgagcaagatgtagtcaacacacacacacacacacac 2580  
QY 2581 ccagccaggaatcaaggttctgcacccagtgatctccagagtgagaaacatccccaagaat 2640  
|||||  
Db 2581 ccagccaggaatcaaggttctgcacccagtgatctccagagtgagaaacatccccaagaat 2640  
QY 2641 ggaacttggagagatgtctcaatcgttgagatacagagcttcttgacagggccaacaac 2700  
|||||  
Db 2641 ggaacttggagagatgtctcaatcgttgagatacagagcttcttgacagggccaacaac 2700  
QY 2701 tcgltgggctagctaaagcaatctgaagcttaataaglatgaacthaaaatgcatgtagt 2760  
|||||

Db 2701 tcgltgggctagctaaagcaatctgaagcttaataaglatgaacthaaaatgcatgtagt 2760  
|||||  
QY 2761 gtaagaactcatgagagacatggaatatgtatctccgacatgtaacagtataaactcga 2820  
|||||  
Db 2761 gtaagaactcatgagagacatggaatatgtatctccgacatgtaacagtataaactcga 2820  
QY 2821 gttccatccacttctctatgaataaacaagagtgtagatataatgaacactc 2880  
|||||  
Db 2821 gttccatccacttctctatgaataaacaagagtgtagatataatgaacactc 2880  
QY 2881 tatgacacttattgtctcatgataaattccctctattatataaatacatctgaactg 2940  
|||||  
Db 2881 tatgacacttattgtctcatgataaattccctctattatataaatacatctgaactg 2940  
QY 2941 acggctttagaagcttccaatagtaacaaaacaatgtgtactataagaacttctaa 3000  
|||||  
Db 2941 acggctttagaagcttccaatagtaacaaaacaatgtgtactataagaacttctaa 3000  
QY 3001 caattctaactttagcatgtggaacgagacataagtgtagaagaagacataacaattaa 3060  
|||||  
Db 3001 caattctaactttagcatgtggaacgagacataagtgtagaagaagacataacaattaa 3060  
QY 3061 tggagaaggttgtctccattatataatataatataccactttagtataataga 3120  
|||||  
Db 3061 tggagaaggttgtctccattatataatataatataccactttagtataataga 3120  
QY 3121 tgttaaggagacataacaatatataaagagaagttgtatccattatataatataac 3180  
|||||  
Db 3121 tgttaaggagacataacaatatataaagagaagttgtatccattatataatataac 3180  
QY 3181 taacctattatataatataactataccacttattatgctcttataaagtttagacatg 3240  
|||||  
Db 3181 taacctattatataatataactataccacttattatgctcttataaagtttagacatg 3240  
QY 3241 atatttctaataatttagtgatataatgataatgaaaggttactatttgaactcttactac 3300  
|||||  
Db 3241 atatttctaataatttagtgatataatgataatgaaaggttactatttgaactcttactac 3300  
QY 3301 tgtataaaggtttagatcatctcttaagtggtctatattatattatgtcttcttaacat 3360  
|||||  
Db 3301 tgtataaaggtttagatcatctcttaagtggtctctatattatattatgtcttcttaacat 3360  
QY 3361 aaaaaaaaatctatagattgtgttgatataaataatgaaaggtttaaaataataaata 3420  
|||||  
Db 3361 aaaaaaaaatctatagattgtgttgatataaataatgaaaggtttaaaataataaata 3420  
QY 3421 ataaataacataataatataatgataaataatattatataataaaacatttatctataaaa 3480  
|||||  
Db 3421 ataaataacataataatataatgataaataatattatataataaaacatttatctataaaa 3480  
QY 3481 agtaaatatgtcatataaatacataacatcgctttagccttgcgtgagacactcaatatt 3540  
|||||  
Db 3481 agtaaatatgtcatataaatacataacatcgctttagccttgcgtgagacactcaatatt 3540  
QY 3541 taaacgagagtaaacataattggaactttagtattttaacaatatattatgaacact 3600  
|||||  
Db 3541 taaacgagagtaaacataattggaactttagtattttaacaatatattatgaacact 3600  
QY 3601 atgaaatttttttttatctgcgcaaggaataaataatgaagaggaacatgtgt 3660  
|||||  
Db 3601 atgaaatttttttttatctgcgcaaggaataaataatgaagaggaacatgtgt 3660  
QY 3661 gtcccaatccctatacaacaacacttccacaggaagtgtagtgcgggagacacaaaanaac 3720  
|||||  
Db 3661 gtcccaatccctatacaacaacacttccacaggaagtgtagtgcgggagacacaaaanaac 3720  
QY 3721 aggcagaaggaatttttaattgt 3780  
|||||  
Db 3721 aggcagaaggaatttttaattgt 3780  
QY 3781 caactacataaacctttagcagtagagcaatgtgtgacggtgtgctttagcttcttta 3840  
|||||

Db 3781 cactacataaacctttagcagtagagcaatggttgcacgtgtgcttagcttcttla 3840  
Qy 3841 ttatttttttaccacaagaataaataaataatagacacttccagatgttcc 3900  
Db 3841 ttatttttttaccacaagaataaataaataatagacacttccagatgttcc 3900  
Qy 3901 aaccttatacaaaacccaagaagttctccctagacacccaactaagtacc 3957  
Db 3901 aaccttatacaaaacccaagaagttctccctagacacccaactaagtacc 3957  
RESULT 2  
AAZ10376  
ID AAZ10376 standard; DNA: 1558 BP.  
XX  
AC AAZ10376;  
XX  
DT 15-NOV-1999 (first entry)  
XX  
DE Nucleotide sequence of the bean phaseolin promoter.  
XX  
KW Eco gene: plant metabolism; fatty acid oxidation enzyme;  
KW polyhydroxyalkanoate; oil composition; seed production; plant biomass;  
KW transgenic plant; promoter; bean; ss.  
XX  
OS Phaseolus sp.  
XX  
PN W09945122-A1.  
XX  
PD 10-SEP-1999.  
XX  
PF 05-MAR-1999; 99WO-US04999.  
XX  
PR 06-MAR-1998; 98US-0077107.  
XX  
PA (META-) METABOLIX INC.  
XX  
PI Boynton L, Hulsman GW, Moloney M, Patterson N, Peoples OP;  
PI Snell K;  
XX  
DR WPI: 1999-540850/45.  
XX  
PT Modifying fatty acid metabolism in plants, useful for increasing  
PT biomass and producing specific polymers in seeds  
XX  
PS Example 3; Page 63; 79pp; English.  
XX  
CC The present sequence represents the bean phaseolin promoter, which  
CC is used to construct plasmids for the expression of the P. putida  
CC faoA gene. The gene encodes an enzyme that may be used in the  
CC method of the invention. The specification describes a method for  
CC manipulating the metabolism of a plant, and comprises expressing a  
CC heterologous gene encoding fatty acid oxidation enzymes in the  
CC cytosol or plastids other than the peroxisomes, glyoxisomes or  
CC mitochondria of the plant. The method may be used to enhance the  
CC biological production of polyhydroxyalkanoates or novel oil compositions  
CC in a transgenic plant. Plants which may be used to produce these  
CC compounds in this way include Brassicas, maize, soybean, cottonseed,  
CC sunflower, palm, coconut, safflower, peanut, mustards, flax, tobacco and  
CC alfalfa. The method may also be used to prevent or suppress seed  
CC production and therefore increase the production of biomass (leaves,  
CC stems, stalks) by plants.  
XX  
SQ Sequence 1558 BP; 516 A; 310 C; 171 G; 561 T; 0 other;

Query Match 39.0%; Score 1544; DB 20; Length 1558;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1544; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 catgtactccagatataatagatgttggctctcgcgcggtgtttttac 71  
Db 8 catgtactccagatataatagatgttggctctcgcgcggtgtttttac 67

Qy 72 ctctattaaagggtttccacctaanaattcgtgatactctcacttactgttac 131  
Db 68 ctctattaaagggtttccacctaanaattcgtgatactctcacttactgttac 127  
Qy 132 ttaatttctcaatacttggttgaaatatacaagcttcgcacagataccctcaaa 191  
Db 128 ttaatttctcaatacttggttgaaatatacaagcttcgcacagataccctcaaa 187  
Qy 192 atttattattgttaaacattttcaaacgcgataaattttatgaagttccgctactct 251  
Db 188 atttattattgttaaacattttcaaacgcgataaattttatgaagttccgctactct 247  
Qy 252 taatgtagtcaacattttcatalatgaatataataatttacttaattttagcgttggtag 311  
Db 248 taatgtagtcaacattttcatalatgaatataataatttacttactttagcgttggtag 307  
Qy 312 aaagcataaagaattatcttctctctctcctcctaataatgtttatatacataataaca 371  
Db 308 aaagcataaagaattatcttctctctcctcctaataatgtttatatacataataaca 367  
Qy 372 aatctttaccttgaagaagattcccatattatattttaaataatataataat 431  
Db 368 aatctttaccttgaagaagattcccatattatattttaaataatataataat 427  
Qy 432 ttccaacacgtaaatctcataataaagttgtttccaagaatataaatttaactcc 491  
Db 428 ttccaacacgtaaatctcataataaagttgtttccaagaatataaatttaactcc 487  
Qy 492 ataattttttatctcgcagatctctaaagcaacccagtgacacactacgttttt 551  
Db 488 ataattttttatctcgcagatctctaaagcaacccagtgacacactacgttttt 547  
Qy 552 tctttgaaataaaaaaattccaattatcatgtattttttttatacaagaatattcca 611  
Db 548 tctttgaaataaaaaaattccaattatcatgtattttttttatacaagaatattcca 607  
Qy 612 aacaatcatttggtaattcttgcgaagcaagtcagtgttaagcaaatctcataattccat 671  
Db 608 aacaatcatttggtaattcttgcgaagcaagtcagtgttaagcaaatctcataattccat 667  
Qy 672 ttgacacacggaagtaactgaagatctgcttttaacatgagagacactcttcaagta 731  
Db 668 ttgacacacggaagtaactgaagatctgcttttaacatgagagacactcttcaagta 727  
Qy 732 attttaataagttacatataatccaagttccataatacaataactcaataacttctca 791  
Db 728 attttaataagttacatataatccaagttccataatacaataactcaataacttctca 787  
Qy 792 aaaaattaatagatataataaataattactttttaatttaagttaattgttgaat 851  
Db 788 aaaaattaatagatataataaataattactttttaatttaagttaattgttgaat 847  
Qy 852 ttgagacattgatttatattactactatggtttaattggtttatagatatttaagta 911  
Db 848 ttgagacattgatttatattactactatggtttaattggtttatagatatttaagta 907  
Qy 912 aataaagtaagtgaatgaatggttagagttgtaaccttaaacataaactataacatta 971  
Db 908 aataaagtaagtgaatgaatggttagagttgtaaccttaaacataaactataacatta 967  
Qy 972 ttgtgactaatttcatatatttcttatttgccttacttcttctgtgatagttcg 1031  
Db 968 ttgtgactaatttcatatatttcttatttgccttacttcttctgtgatagttcg 1027  
Qy 1032 taactagaatacagttggttgccaatggcactcgtgtgctttgtgttcacatgcatggc 1091  
Db 1028 taactagaatacagttggttgccaatggcactcgtgtgctttgtgttcacatgcatggc 1087  
Qy 1092 ttgcgcaagaagaagaacaagaagaagaagaagaagaagaagaagaagaagaagaaga 1151  
Db 1088 ttgcgcaagaagaagaacaagaagaagaagaagaagaagaagaagaagaagaagaaga 1147



Db	907	aataaagtaatgtagtagagtgtagagtgtaagtgtaacccaataaacaataaactaagattta	966
Qy	972	tgggtgagcctaatttcataabatttccttattgcttacccttctctggtatgtaagtcg	10311
Db	967	tgggtgagcctaatttcataabatttccttattgcttacccttctctggtatgtaagtcg	10266
Qy	1032	taactaagaaattacaagtggtgtgcacagagcactctgtgtgtcttcttggttcagatgggct	10921
Db	1027	taactaagaaattacaagtggtgtgtgcacagagcactctgtgtgtcttcttggttcagatgggct	10866
Qy	1092	tttgcgcaagaaaaagacaagaacacaagaaaaagacaagaacagagagacaacgcaat	11521
Db	1087	tttgcgcaagaaaaagacaagaacacaagaaaaagacaagaacagagagacaacgcaat	11466
Qy	1152	cacacaaccaactcaaatattagtcactgtgcgtgtaaccaagatcgcgcgccatgtaagtcta	12111
Db	1147	cacacaaccaactcaaatattagtcactgtgcgtgtaaccaagatcgcgcgccatgtaagtcta	12066
Qy	1212	aatgcatgccaagcaaacacagctgtctaactgtacatttaatggttcacccatctcaac	12711
Db	1207	aatgcatgccaagcaaacacagctgtctaactgtacatttaatggttcacccatctcaac	12666
Qy	1272	cacacacaacaacagctgtcctttcttcatacatcacacaacacotgtatatattcaat	13311
Db	1267	cacacacaacaacagctgtcctttcttcatacatcacacaacacotgtatatattcaat	13266
Qy	1332	ctcttcgcgcacactcaattttcttctactcttcacaacacgltcaactgtcatatgctgtcatc	13911
Db	1327	ctcttcgcgcacactcaattttcttctactcttcacaacacgltcaactgtcatatgctgtcatc	13866
Qy	1392	ccatgcccacaatctccatgtatgattccaaacacactctctcttataataataactataat	14511
Db	1387	ccatgcccacaatctccatgtatgattccaaacacactctctcttataataataactataat	14466
Qy	1452	acccttaatatcactactctcttcatacatcacatccatccagagtaactactactacta	15111
Db	1447	acccttaatatcactactctcttcatacatcacatccatccagagtaactactactacta	15066
Qy	1512	ctataatacccccaaccacactcatatccaatactactac	1552
Db	1507	ctataatacccccaaccacactcatatccaatactactac	1547
RESULT 4			
AAZ10392			
ID	AAZ10392	standard; DNA; 1244 BP.	
XX	AAZ10392;		
AC			
XX			
DT	15-NOV-1999	(first entry)	
XX			
DE	Nucleotide sequence of the bean phaseolin terminator.		
XX			
XX			
OS	Phaseolus sp.		
XX			
PN	W09945122-A1.		
XX			
PD	10-SEP-1999.		
XX			
PF	05-MAR-1999;	99WC-US04999.	
PR	06-MAR-1998;	98US-0077107.	
XX			
PA	(META-) METABOLIX INC.		
XX			
PI	Boynton L, Huisman GW, Moloney M, Patterson N, Peoples OP;		
XX	Snell K;		
DR	WPI; 1999-540850/45.		

XX Modifying fatty acid metabolism in plants, useful for increasing  
PT biomass and producing specific polymers in seeds  
XX  
PS  
XX  
XX Example 3; page 71; 79pp; English.  
CC  
CC The present sequence represents the bean phaseolin terminator, which  
CC is used to construct plasmids for the expression of the P. putida  
CC fAOAB gene. The gene encodes an enzyme that may be used in the  
CC method of the invention. The specification describes a method for  
CC manipulating the metabolism of a plant, and comprises expressing a  
CC heterologous gene encoding fatty acid oxidation enzymes in the  
CC cytosol or plastids other than the peroxisomes, glyoxisomes or  
CC mitochondria of the plant. The method may be used to enhance the  
CC biological production of polyhydroxyalkanoates or novel oil compositions  
CC in a transgenic plant. Plants which may be used to produce these  
CC compounds in this way include Brassicas, maize, soybean, cottonseed  
CC sunflower, palm, coconut, safflower, peanut, mustards, flax, tobacco  
CC alfalfa. The method may also be used to prevent or suppress seed  
CC production and therefore increase the production of biomass (leaves,  
CC stems, stalks) by plants.

Sequence 1244 BP; 456 A; 174 C; 178 G; 436 T; 0 other;

Query Match	31.0%	Score 1226;	DB 20;	Length 1244;
Best Local Similarity	100.0%	Pred. No. 0;		
Matches 126; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0;

QY	2732	aataagatctgaacctaanaatgcgtagtggtgttaagaactatggaagagcaatggaatattggt	2739
Db	19	ataagatctgaacctaanaatgcgtagtggtgttaagaactatggaagagcaatggaatattggt	78
QY	2792	atccgacacatgtaacgaatataataactgagctccgactccatctcattctctatgataaacaa	2851
Db	79	atccgacacatgtaacgaatataataactgagctccatctcattctctatgataaacaa	138
QY	2852	aggatgttaagatataatataaacctctatactgacacctatgttctatgataaatttcc	2911
Db	139	aggatgtatgatataatataaacacctaactgacacctatgttctatgataaatttcc	198
QY	2912	tctttatcttaataatcaatcgatctgtagcggtctatggaatgcttcaaatagtacaaa	2977
Db	199	tctttatcttaataatcaatctcgatctgtagcggtctatggaatgcttcaaatagtacaaa	258
QY	2972	aacaaatgtgtactataagaacttcttaaacaaacttaacttaagcatgtgtgaagagaca	3033
Db	259	aacaaatgtgtactataagaacttcttaaacaaacttaactttagcatgtgtgaagagaca	318
QY	3032	taagtggttaagaagaacataaacatataatgaaggaagaagttgtctccattatattat	3091
Db	319	taagtggttaagaagaacataaacatataatgaaggaagaagttgtctccattatattat	378
QY	3092	atataccacacttctgtaattatataatlaagatggttaaggaagacataaacattataagaag	3155
Db	379	atataccacacttctgtaattatataatlaagatggttaaggaagacataaacattataagaag	438
QY	3152	aagtttgatccatttatataatatataacacccatttatatatattatatttccaacta	3211
Db	439	aagtttgatccatttatataatatataacacccatttatatatattatatttccaacta	498
QY	3212	tttaagtctctataaagtttgataccatgtaattcttcaattattttagtctgatagtatata	3271
Db	499	tttaagtctctataaagtttgataccatgtaattcttcaattattttagtctgatagtatata	558
QY	3272	gaaagagtgatactattgaaactctctactctgtataaaggttggaatcaaccttaagaatg	3331
Db	559	gaaagagtgatactattgaaactctctactctgtataaaggttggaatcaaccttaagaatg	618
QY	3332	tctatttaatttttggctctctacgaataaaaaaaatatgagtttggtttgataaaaa	3391
Db	619	tctatttaatttttggctctctacgaataaaaaaaatatgagtttggtttgataaaaa	678



Oy	3392	tatgaagagatttaataataataataataataacatatataatagtataataat	3451
Db	679	tatgaaggattttaaataataataataataacatatataatagtataataat	738
Oy	3452	tattataataaacattactctataaaagaataattgtccaatcattacaatcgt	3511
Db	739	tattataataataacattactctataaaagaataattgtccaatcattacaatcgt	798
Oy	3512	ttagccttgctgagcagctcattcaattttaacagaglaaacaatttgtacttttg	3571
Db	799	ttagccttgctgagcagctcattcaattttaacagaglaaacaatttgtacttttg	858
Oy	3572	ttatttaacaaaattattatttaacacatatgaattttttttttlalcggcaagaa	3631
Db	859	ttatttaacaaaattattatttaacacatatgaattttttttttlalcggcaagaa	918
Oy	3632	taaaattaatitagagagcaatlggtgtcccaatccttatacaaccaattccaag	3691
Db	919	taaaattcaaatlagagagcaatlggtgtcccaatccttatacaaccaattccaag	978
Oy	3692	gaagtcaggtcgcggggcacaaaaaacagcgaaggaaatttttaatttggtgtc	3751
Db	979	gaagtcaggtcgcggggcacaaaaaacagcgaaggaaatttttttaatttggtgtc	1038
Oy	3752	tgtgttcctgcataaatattatgatcaglaaanaacactacacataaccctttagcsgtagca	3811
Db	1039	tgtgttcctgcataaatattatgatcaglaaanaacactacacataaccctttagcsgtagca	1098
Oy	3812	atggtttgacctgtgccttaactcttttattttttttttttlalcagcaagaataata	3871
Db	1099	atggtttgacctgtgccttaactcttttattttttttttttlalcagcaagaataata	1158
Oy	3872	aataaatgagaccccttcagagatgtttcacccctatacaaaaccccaaaacagtt	3931
Db	1159	aataaatgagaccccttcagagatgtttcacccctatacaaaaccccaaaacagtt	1218
Oy	3932	cctagcacccctaccactaagtagcc	3957
Db	1219	cctagcacccctaccactaagtagcc	1244
<hr/>			
RESULT	5		
AAS00569			
ID	AAS00569	standard; DNA; 1173 BP.	
XX	XX		
AC	AAS00569;		
XX	XX		
DT	14-MAY-2001	(first entry)	
XX	XX		
DE	Bovine pre-pro-chymosin DNA sequence.		
XX	XX		
KM	Chymosin; transcription regulator; terminator sequence; soybean; corn;		
KM	pre-pro-chymosin; rape seed; sunflower; cotton; tobacco; alfalfa; wheat;		
KM	baleyle; oats; sorghum; Arabidopsis thaliana; potato; flax; linseed; rice;		
KM	safflower; oil palm; ground nut; Brazil nut; coconut; castor; coriander;		
KM	squash; jojoba; ds.		
XX	OS		
XX	Bos sp.		
XX	XX		
FH	Key	Location/Qualifiers	
FT	CDS	1..1173	
FT	FT	/tag- a	
FT	FT	/product= "Bovine chymosin"	
FT	FT	1..78	
FT	FT	/tag- b	
FT	FT	79..201	
FT	FT	/tag- c	
FT	FT	/note= "Pro sequence"	
FT	FT	mat_peptide	
FT	FT	202..1170	
FT	FT	/tag- d	
FT	FT	/product= "Mature bovine chymosin"	
XX	XX		
NN	MO200114571-A1.		

Query Match	29.6%	Score 1173:	DB 22:	Length 1173:
Best Local Similarity	100.0%	Pred. No. 0:		
Matches 1173:	Conservative	0:	Mismatches	0:
			Indels	0:
			Gaps	0:
1554 atgaacttccttaagtcttccttccttccttaagcttccttggttcgtgaataactcgtt	1613			
Db 1 atgaacttccttaagtcttccttccttccttccttccttccttcgttcgtgaataactcgtt	60			
Qy 1614 gctgttactcaagcgcgtctgtagatacaaccgcatactctctcaaaaagtaagcttcggt	1673			
Db 61 gctgttactcaagcgcgtctgtagatacaaccgcatactctctcaaaaagtaagcttcggt	120			
Qy 1674 aagcgcgctgaaggaacatgagctcttagaagaactctctgcagaaaacaacagtaigcacc	1733			
Db 121 aagcgcgctgaaggaacatgagctcttagaagaactctctgcagaaaacaacagtaigcacc	180			
Qy 1734 agcagcaagtaactccggtctcggtgaagtctgtaagctgtaagcgtaccataactacttgat	1793			
Db 181 agcagcaagtaactccggtctcggtgaagtctgtaagctgtaagcgtaccataactacttgat	240			
Qy 1794 agtcaatacttgggaaagatctcaactccggaaccccgcccaagagttcaacggttccttt	1853			
Db 241 agtcaatacttgggaaagatctcaactccggaaccccgcccaagagttcaacggttccttt	300			
Qy 1854 gatactggttcctcgtgaactctgggttcctctatactatactgaagaagcaatgcgtgcaag	1913			
Db 301 gatactggttcctcgtgaactctgggttcctctatactatactgaagaagcaatgcgtgcaag	360			
Qy 1914 aaccacccaagttcgatccgcgaagaagtcgtccaccttcacagaactttggcaaaccttg	1973			
Db 361 aaccacccaagttcgatccgcgaagaagtcgtccaccttcacagaactttggcaaaccttg	420			
Qy 1974 tctatacaataagtagacagtagaactgaagaagtaactttaggtcgtatccgctgaactg	2033			
Db 421 tctatacaataagtagacagtagaactgaagaagtaactttaggtcgtatccgctgaactg	480			
Qy 2034 tccaacatgttgacattccaacagacagtaggacttagacaccaagaacagtgatgctc	2093			
Db 481 tccaacatgttgacattccaacagacagtaggacttagacaccaagaacagtgatgctc	540			

QY 2094 ttccactatgcagaattcgaatcgtgcatccttggtatgtgcatacccatcgctcgctcagag 2153  
|||||  
Db 541 ttccactatgcagaattcgaatcgtgcatccttggtatgtgcatacccatcgctcgctcagag 600  
QY 2154 taactgcatacctgtgtttgcacaatgatgaacccagacactagatagctcaagactgttc 2213  
|||||  
Db 601 ttaccgtatacctgtgtttgcacaatgatgaacccagacactagatagctcaagactgttc 660  
QY 2214 tcggtttacatgcagcagaatgcgcagcagagacatgctcaccgcttgcagcttatgaccca 2273  
|||||  
Db 661 tcggtttacatgcagcagaatgcgcagcagagacatgctcaccgcttgcagcttatgaccca 720  
QY 2274 tccctactacacagagatcctcttcaactggttccagttcactgctgcagcagctatgagcaattc 2333  
|||||  
Db 721 tccctactacacagagatcctcttcaactggttccagttcactgctgcagcagctatgagcaattc 780  
QY 2334 actgtgcagcagtgctcaccatcagcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2393  
|||||  
Db 781 actgtgcagcagtgctcaccatcagcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 840  
QY 2394 atcttgataccggttgcagcagcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2453  
|||||  
Db 841 atcttgataccggttgcagcagcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 900  
QY 2454 caagctatgtgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2513  
|||||  
Db 901 caagctatgtgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 960  
QY 2514 agctatagctcctacagctgt 2573  
|||||  
Db 961 agctatagctcctacagctgt 1020  
QY 2574 ggcctataccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2633  
|||||  
Db 1021 ggcctataccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1080  
QY 2634 cagaatgtgactgtgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2693  
|||||  
Db 1081 cagaatgtgactgtgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1140  
QY 2694 aacaacctcggttggt 2726  
|||||  
Db 1141 aacaacctcggttggt 1173

RESULT 6  
AAD17528  
ID AAD17528 standard; DNA; 2970 BP.  
XX  
AC AAD17528;  
XX  
DT 10-DEC-2001 (first entry)  
XX  
DE Kunitz soybean trypsin inhibitor (KSTI) encoding DNA.  
XX  
KW Soybean; allergen; transgenic plant; P34 protein; Gly m Bd 30K; Gly m 1;  
KW soybean vacuolar protein; Gly m IA; Gly m IB; RGLY m3; Glycinin G1;  
KW alablb; food; infant formula; animal feed; coating; salad oil; syrup;  
KW spraying oil; roasting oil; frying oil; cracker; confectionery product;  
KW snack food; topping; sauce; batter; breeding mixture; baking mix; dough;  
KW Kunitz soybean trypsin inhibitor; KSTI; KTI3; ds.  
XX  
OS Glycine max.  
XX  
PN WO200168867-A2.  
XX  
PD 20-SEP-2001.  
XX  
PE 15-MAR-2001; 2001WO-US08254.  
XX  
PR 16-MAR-2000; 2000US-0189823.  
XX

PA (DUPO) DU PONT DE NEMOURS & CO E I.  
PA (PION-) PIONEER HI-BRED INT INC.  
XX  
PI Jung R, Kinney AJ;  
XX  
DR WPI: 2001-582460/65.  
XX  
PT Recombinant expression construct to lower allergen (e.g. Gly m Bd 30K)  
PT content of a soybean, comprises a nucleic acid fragment encoding the  
PT allergen, useful for producing soybean plants which can be used to make  
PT soybean products  
XX  
PS Claim 4; Page 48-49; 57pp; English.  
XX  
CC The patent discloses hypoallergenic transgenic soybeans and recombinant  
CC expression constructs to lower soybean vacuolar protein, commonly known  
CC as P34 (Gly m Bd 30K or Gly m 1) and other allergens such as Gly m IA,  
CC Gly m IB, RGLY m3 or Glycinin G1 (alablb). The allergen content of the  
CC soybean is reduced by sense suppression which is accomplished by using  
CC the expression construct that comprises a nucleic acid fragment encoding  
CC the allergen. The constructs are useful for producing hypoallergenic  
CC transgenic soybean plants which can be used to make hypoallergenic  
CC soybean products which can be used in a variety of food (e.g. infant  
CC formulas) and animal feed applications. The oil made from seeds of the  
CC hypoallergenic transgenic soybean plants can be used as ingredients,  
CC as coatings, as salad oils, as spraying oils, as roasting oils, and  
CC as frying oils. The foods in which the oil may be used include crackers  
CC and snack foods, confectionery products, syrups and toppings, sauces,  
CC batter and bread mixtures, baking mixes and doughs. The present  
CC sequence is a DNA encoding kunitz soybean trypsin inhibitor (KSTI  
CC or KTI3), a minor soybean seed allergen.  
CC  
SQ Sequence 2970 BP; 1018 A; 547 C; 527 G; 878 T; 0 other;

Query Match 22.4%; Score 887; DB 22; Length 2970;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 887; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2735 aagatgacctaataatgcatgtgtaagagcgtcatgagcagatggaatatgtatc 2794  
|||||  
Db 1796 aagatgacctaataatgcatgtgtaagagcgtcatgagcagatggaatatgtatc 1855  
QY 2795 cgacatgtaacagbataataacgtgagctccatctcactcttctatgaatacaaaagg 2854  
|||||  
Db 1856 cgacatgtaacagbataataacgtgagctccatctcactcttctatgaatacaaaagg 1915  
QY 2855 atgtatgatataataaacctctatctatgcaccttatgtgtcatgataattcctct 2914  
|||||  
Db 1916 atgtatgatataataaacctctatctatgcaccttatgtgtcatgataattcctct 1975  
QY 2915 tatatataatcatctgcatcgtgacgcttatgagatgcttcaaatgatacaaac 2974  
|||||  
Db 1976 tatatataatcatctgcatcgtgacgcttatgagatgcttcaaatgatacaaac 2035  
QY 2975 aaatgttactataagacttcttaacaactcttaactttagcatgtgagcagagacataa 3034  
|||||  
Db 2036 aaatgttactataagacttcttaacaactcttaactttagcatgtgagcagagacataa 2095  
QY 3035 gtgttaagaagaacataaatatataatgaagaagttgtgttcatttatatatata 3094  
|||||  
Db 2096 gtgttaagaagaacataaatatataatgaagaagttgtgttcatttatatatata 2155  
QY 3095 ttaccacttatgtatatataatgaagatgtaagagacatacaaatataagaagagag 3154  
|||||  
Db 2156 ttaccacttatgtatatataatgaagatgtaagagacatacaaatataagaagagag 2215  
QY 3155 ttgtatcatttatatatataatactaacatttatatatataataccacttatt 3214  
|||||  
Db 2216 ttgtatcatttatatatataatactaacatttatatatataataccacttatt 2275  
QY 3215 aatgtcttataaggtttgcatccatgatattcttaatatatttagttgatatgatagaa 3274  
|||||

```
Db 2276 aatgctttaaaggcttgatccatgataattctaatattttagttgatatglatgaa 2335
QY 3275 agggactatttgaaactctctactctgtataaagttggatcatccttaagtggtct 3334
    |||||||
Db 2336 agggactatttgaaactctctactctgtataaagttggatcatccttaagtggtct 2395
QY 3335 attaatcttctgtctctcaagaataaaaaaatatgagttggttctgataaatat 3394
    |||||||
Db 2396 attaatcttctgtctctcaagaataaaaaaatatgagttggttctgataaatat 2455
QY 3395 tgaagagatttaaaataataataataataataataataataataataataat 3454
    |||||||
Db 2456 tgaagagatttaaaataataataataataataataataataataataataat 2515
QY 3455 tataataaacattatctataaataaagtaattgtcatataactatcacatcgctta 3514
    |||||||
Db 2516 tataataaacattatctataaataaagtaattgtcatataactatcacatcgctta 2575
QY 3515 gccttgcgcgacgactccaattttaaagcagagataaacatattgacttttgta 3574
    |||||||
Db 2576 gccttgcgcgacgactccaattttaaagcagagataaacatattgacttttgta 2635
QY 3575 tttaacaattatatttaaacactatagaaatttttttttttc 3621
    |||||||
Db 2636 tttaacaattatatttaaacactatagaaatttttttttttc 2682
```

## RESULT 7

```
AAAN30063
ID AAAN30063 standard; cDNA; 1098 BP.
XX
AC AAAN30063;
XX
DF 14-JUN-1992 (first entry)
XX
DE Sequence of prorennin cDNA in PCR 10001.
XX
KW Rennin; renin; enzyme; protease; ss.
XX
OS Bos taurus.
XX
FH Key Location/Qualifiers
FT CDS 1..1098
FT /*tag= a
XX
PN EP73029-A.
XX
PD 02-MAR-1983.
XX
PF 19-AUG-1982; 82EP-0107601.
XX
PR 24-AUG-1981; 81JP-0131631.
XX
PA (BEPP/) BEPPU T.
XX
PI Beppu T, Uozumi T, Nishimori K;
XX
DR WPI: 1983-22976K/10.
DR P-PSDB; AAP30603.
XX
PT Plasmid contg. calf pro:rennin DNA - and transformed
PT microorganisms
XX
PS Example; Page 20-23; 32pp; English.
XX
CC The inventors claim recombinant plasmids contg. the cDNA of calf
CC prorennin. Specified plasmids are PCR 10001 and PCR2001 (contg. the
CC whole sequence plus the lac promoter region). Also new are
CC microorganisms transformed with the plasmids, esp. E. coli CRI (ATCC
CC 391710) contg. plasmid PCR2001.
XX
SQ Sequence 1098 BP; 252 A; 326 C; 301 G; 219 T; 0 other;
```

```
Query Match 1.2%; Score 49; DB 4; Length 1098;
Best Local Similarity 100.0%; Pred. No. 1.5e-11;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

## RESULT 8

```
AAQ04683
ID AAQ04683 standard; DNA; 1098 BP.
XX
AC AAQ04683;
XX
DT 05-OCT-1990 (first entry)
XX
DE Sequence encoding calf pro-rennin.
XX
KW Pro-rennin; ds.
XX
OS Bos taurus.
XX
PN JP02109984-A.
XX
PD 23-APR-1990.
XX
PF 01-JAN-1988; 88JP-0302176.
XX
PR 01-JAN-1988; 88JP-0302176.
XX
PA (BEPP/) BEPPU T.
XX
DR WPI: 1990-168358/22.
DR P-PSDB; AAR05080.
XX
PT Complex plasmid and microbe - contains calf pro-rennin cDNA.
PS Disclosure: 32; 13pp; Japanese.
XX
CC Protein product may be expressed in E.coli expression system from
CC plasmid pBR322.
XX
SQ Sequence 1098 BP; 252 A; 327 C; 300 G; 219 T; 0 other;
```

```
Query Match 1.2%; Score 49; DB 11; Length 1098;
Best Local Similarity 100.0%; Pred. No. 1.5e-11;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 2209 tgtctcggttacatgacaggaatggccagagagatgctcacgct 2257
    |||||||
Db 581 tgtctcggttacatgacaggaatggccagagagatgctcacgct 629
```

## RESULT 9

```
AAAN40295
ID AAAN40295 standard; mRNA; 1175 BP.
XX
AC AAAN40295;
XX
DT 04-FEB-1992 (first entry)
XX
DE Sequence encoding a polypeptide displaying milk clotting activity.
XX
KW Cheese-making; recombinant protein; rennet substitute; milk clot; ss.
XX
OS Bos taurus.
XX
FH Key Location/Qualifiers
FT sig_peptide 21..69
FT /*tag= a
```



XX Recombinant DNA constructs for expressing protein in milk -  
PT contg. specific mammary gland transcription control region and  
PT signal sequence, providing high yield and easy prod. recovery  
XX  
PS Disclosure; Page 21; 41pp; German.  
XX  
CC The rennin gene was isolated using the probe represented in AAQ14775.  
CC It was used as heterologous peptide/protein together with parts of the  
CC alpha-SI-casein gene in the prodn. of DNA constructs. The heterologous  
CC peptide or protein may also be human insulin-like growth factor I.  
CC The constructs provide high yields of the protein with simple recovery  
CC from the milk. Activation of the gene occurs only in the mammary gland.  
CC See also AAQ14050, AAQ14774-77.  
XX  
SQ Sequence 1210 BP; 278 A; 356 C; 324 G; 252 T; 0 other;

Query Match 1.2%; Score 49; DB 12; Length 1210;  
Best Local Similarity 100.0%; Pred. No. 1.4e-11;  
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2209 tgtctcgtttacatgacaggaatgcccagagagatctcagct 2257  
|||||  
Db 643 tgtctcgtttacatgacaggaatgcccagagagatctcagct 691

RESULT 12  
AAAT03006  
ID AAAT03006 standard; DNA; 1240 BP.  
XX  
AC AAAT03006;  
XX  
DT 13-JUN-1996 (first entry)  
XX  
DE Chymosin open reading frame.  
XX  
KM al-3; albino mutant; light-regulated; Neurospora; bread mould;  
KM heterologous gene; expression; control; chymosin; ss.  
XX  
OS Mammalian sp.  
XX  
PN WO9530739-A1.  
XX  
PD 16-NOV-1995.  
XX  
PF 09-MAY-1995; 95MO-US05716.  
XX  
PR 10-MAY-1994; 94US-0240372.  
XX  
PA (UYHA-) UNIV HAWAII.  
XX  
PI Kato EK, Stuart WD;  
XX  
DR WPI: 1995-404108/51.  
XX  
PT Nucleic acid for expression of heterologous protein - contains  
PT albino promoter for light induced expression in filamentous fungi  
XX  
PS Example 2; Fig 7; 26pp; English.  
XX  
CC The DNA is that of a mammalian gene (open reading frame) encoding  
CC chymosin. The gene was placed in operable linkage with the al-3  
CC promoter (see AAT03005). The al-3 gene controls the production of  
CC geranyl geranyl pyrophosphatase (GGPP) synthetase. GGPP is a precursor  
CC for carotenoids and xanthophylls. It has been shown that exposure to  
CC light increases the transcription level of GGPP synthetase 15-45 fold.  
CC Light activates a number of genes in the common bread mould, Neurospora.  
CC This can be used to regulate the expression of genes encoding  
CC heterologous proteins, e.g. chymosin, in recombinant production systems.  
CC Use of a light-regulated promoter is a simple and effective way to  
CC control expression and allows timing to be adapted to the physiological  
CC status of the host.

XX  
SQ Sequence 1240 BP; 274 A; 374 C; 339 G; 253 T; 0 other;

Query Match 1.2%; Score 49; DB 16; Length 1240;  
Best Local Similarity 100.0%; Pred. No. 1.4e-11;  
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2209 tgtctcgtttacatgacaggaatgcccagagagatctcagct 2257  
|||||  
Db 700 tgtctcgtttacatgacaggaatgcccagagagatctcagct 748

RESULT 13  
AAN40055  
ID AAN40055 standard; DNA; 1278 BP.  
XX  
AC AAN40055;  
XX  
DT 02-FEB-1992 (first entry)  
XX  
DE Sequence of prochymosin gene.  
XX  
KM Prochymosin expression vector; E.coli trp operon; chymosin; ss.  
XX  
OS Bos taurus.  
XX  
FH Key Location/Qualifiers  
FT CDS 3..1130  
FT /\*tag= a  
FT polyA\_signal 1245..1250  
FT /\*tag= b  
XX  
PN EP121775-A.  
XX  
PD 17-OCT-1984.  
XX  
PF 07-MAR-1984; 84EP-0102451.  
XX  
PR 09-MAR-1983; 83JP-0038439.  
XX  
PA (BEPF/) BEPPU T.  
XX  
PI Beppu T, Uozumi T, Nishimori K, Shimizu N, Kawaguchi Y;  
PI Hidaka M;  
XX  
DR WPI: 1984-258001/42.  
DR P-PSDB; AAP40078.  
XX  
PT Expression plasmid comprising prochymosin gene and vector -  
PT useful for transforming Escherichia coli for prochymosin prodn.  
XX  
PS Disclosure; Fig 1; 59pp; English.  
XX  
XX  
CC The inventors claim the prochymosin gene comprising a nucleotide  
CC sequence from (a) the 1st codon (GCT) to the 365th codon (ATC), or  
CC (b) the 5th codon (CGG) to the 365th codon (ATC); and recombinant  
CC plasmids harboured by Escherichia coli strains deposited as FERM BP-  
CC 262, -263 and -264. Any portion of the nucleotide sequence as  
CC described in AAN40055 can be used. Also claimed is a vector derived  
CC from plasmid pBR322. Typically plasmid pCR501 is obtd. from pOCT 2.  
CC The transcriptional direction of pOCT 3 is opposite to that of  
CC pOCT 2; it is clockwise in pOCT 2 (5' to 3') whereas counter  
CC clockwise in pOCT 3.  
XX  
SQ Sequence 1278 BP; 309 A; 384 C; 326 G; 259 T; 0 other;

Query Match 1.2%; Score 49; DB 5; Length 1278;  
Best Local Similarity 100.0%; Pred. No. 1.4e-11;  
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2209 tgtctcgtttacatgacaggaatgcccagagagatctcagct 2257

Db 613 tttctcggttacatgacaggaatgccaagagagacatgtctcaagct 661

## RESULT 14

AAN30022  
ID AAN30022 standard; DNA; 1289 BP.

XX  
AC AAN30022;

XX  
DT 25-APR-1992 (first entry)

XX  
DE Sequence of veal chymosin gene.

XX  
KM Protolytic enzyme; zymogen; rennin; chymosin; cheese making;  
microbial vector; ss.

XX  
OS Bos taurus.

XX  
FH Key Location/Qualifiers

FT CDS 1..1140

XX  
FT /\*tag= a

XX  
PN BE897201-A.

XX  
PD 03-NOV-1983.

XX  
PF 30-JUN-1983; 83BE-0017731.

XX  
PR 13-APR-1983; 83US-0484539.

XX  
PR 01-JUL-1982; 82US-0394433.

XX  
PA (GENE-) GENEX CORP.

XX  
DR WPI: 1983-820813/47.

XX  
DR P-PSDB: AAP30013.

XX  
PS Isolated chymosin or rennin and prochymosin genes - plasmid(s)

XX  
CC from calves, and plasmids contg. the genes which are capable of

XX  
CC replicating in a prokaryotic organism. The prokaryotic organism is

XX  
CC pref. an Escherichia species, esp. E. coli p Gx 1225 (NRRL B-15061).

XX  
CC The microorganisms transformed by the plasmid are also claimed.

XX  
SQ Sequence 1289 BP; 310 A; 391 C; 327 G; 260 T; 1 other;

Query Match 1.2%; Score 49; DB 4; Length 1289;  
Best Local Similarity 100.0%; Pred. No. 1.4e-11;  
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2209 tttctcggttacatgacaggaatgccaagagagacatgtctcaagct 2257  
|||||  
Db 623 tttctcggttacatgacaggaatgccaagagagacatgtctcaagct 671

## RESULT 15

AAN91157  
ID AAN91157 standard; DNA; 1291 BP.

XX  
AC AAN91157;

XX  
DT 07-JUN-1990 (first entry)

XX  
DE Cloned sequence of (pro)chymosin.

XX  
KW Lactic acid bacteria; cheese; Streptococcus cremoris SK112; chymosin;  
proteinnase; PSK112; ss.

XX  
FH Key Location/Qualifiers  
FT precursor\_RNA 29..1123  
FT /\*tag= a  
FT /product=,prochymosin

XX  
PN NL8701378-A.

XX  
PD 02-JAN-1989.

XX  
PE 12-JUN-1987; 87NL-0001378.

XX  
PR 12-JUN-1987; 87NL-0001378.

XX  
PA (NEZU-) NEDERL. INS. ZUIVELON.

XX  
PI Simons AFM, De Vos WM;

XX  
DR WPI: 1989-030097/04.

XX  
DR P-PSDB: AAP94144.

XX  
PT DNA fragment having region specific for lactic acid bacteria -

XX  
PT is contained in plasmid in microorganism used in prodn. of

XX  
PS protein and food prodn. eg cheese.

XX  
PS Disclosure; fig 2; 43pp; Dutch.

XX  
CC The DNA encoding prochymosin can be cloned into a plasmid ( esp. from

XX  
CC S. cremoris SK112) and used to produce large amts of the protein by

XX  
CC recombinant DNA techniques. This could overcome the shortage of

XX  
CC prochymosin due to shortage of calf stomachs and increasing cheese

XX  
CC prodn. Prochymosin is also used in prodn. of yoghurt, butter and

XX  
CC buttermilk. See also AAN91158-N91160.

XX  
SQ Sequence 1291 BP; 323 A; 385 C; 326 G; 257 T; 0 other;

Query Match 1.2%; Score 49; DB 10; Length 1291;  
Best Local Similarity 100.0%; Pred. No. 1.4e-11;  
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2209 tttctcggttacatgacaggaatgccaagagagacatgtctcaagct 2257  
|||||  
Db 609 tttctcggttacatgacaggaatgccaagagagacatgtctcaagct 657

Search completed: August 1, 2002, 01:58:17  
Job time: 16992 sec



**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 23:07:30 ; Search time 5849.81 Seconds  
(without alignments)  
9129.775 Million cell updates/sec

Title: US-09-643-755b-3

Perfect score: 3957  
Sequence: 1 ctgcaggaatcatgtact.....accctacacactaagttacc 3957

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 13736207 seqs, 6748477542 residues

Word size : 0

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

EST:\*  
1: em\_estbda:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hlc:\*  
9: gb\_estl:\*  
10: gb\_est2:\*  
11: gb\_hlc:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	49	1.2	383	10	BG938086
2	49	1.2	430	10	BG937697
3	49	1.2	472	10	BG938320
4	43	1.1	399	10	BG937723
5	25	0.6	240	9	AU073125
6	25	0.6	468	9	AA394733
7	25	0.6	660	12	AG091550
8	24	0.6	136	10	BE763309
9	24	0.6	214	10	C25503
10	24	0.6	312	9	AU053870
11	24	0.6	442	9	AA441444
12	24	0.6	500	9	AU052887
13	24	0.6	560	12	BH366372
14	24	0.6	667	12	AO635511
15	24	0.6	702	10	BG872501
16	24	0.6	727	12	BH522720
17	24	0.6	782	10	BG432457

18	24	0.6	892	10	BG399700	BG399700 602441568
19	23	0.6	178	10	BT974502	BT974502 sat169d07.
20	23	0.6	240	9	AU073229	AU073229 AU073229
21	23	0.6	298	9	BB493759	BB493759 BB493759
22	23	0.6	300	9	BB489847	BB489847 BB489847
23	23	0.6	304	12	AO907383	AO907383 GSSTC0941
24	23	0.6	312	12	BH261085	BH261085 CH230-170
25	23	0.6	325	9	AV532492	AV532492 AV532492
26	23	0.6	339	10	T20903	T20903 2911 Lambda
27	23	0.6	380	9	AV538262	AV538262 AV538262
28	23	0.6	391	9	AA696942	AA696942 OP43d07.s
29	23	0.6	404	9	AA411567	AA411567 zv22g01.s
30	23	0.6	404	12	AO192651	AO192651 HS_2248_B
31	23	0.6	408	9	AV518232	AV518232 AV518232
32	23	0.6	411	9	AV559941	AV559941 AV559941
33	23	0.6	438	9	AA411566	AA411566 zv22g01.r
34	23	0.6	442	12	AO711285	AO711285 HS_5356_B
35	23	0.6	483	9	AV536021	AV536021 AV536021
36	23	0.6	505	9	AI479358	AI479358 tm27e07.x
37	23	0.6	511	9	AV518638	AV518638 AV518638
38	23	0.6	511	9	AV548208	AV548208 AV548208
39	23	0.6	549	9	AV557282	AV557282 AV557282
40	23	0.6	551	12	A2107068	A2107068 RPT-23-2
41	23	0.6	553	9	AV555646	AV555646 AV555646
42	23	0.6	564	12	A2457432	A2457432 1M0260F18
43	23	0.6	565	9	AV532027	AV532027 AV532027
44	23	0.6	574	9	AV542531	AV542531 AV542531
45	23	0.6	581	9	AV439574	AV439574 AV439574

#### ALIGNMENTS

#### RESULT 1

LOCUS BG938086 383 bp mRNA linear EST 11-JUN-2001  
DEFINITION IABO1A08 Bovine Abomasum cDNA Library Bos taurus cDNA 5', mRNA  
ACCESSION BG938086  
VERSION BG938086.1 GI:14337458  
KEYWORDS EST.  
SOURCE COW.  
ORGANISM Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovidae; Bovinae; Bos.  
REFERENCE 1 (bases 1 to 383)  
AUTHORS Moore,S.S., Hansen,C., Li,C., Fu,A., Meng,Y. and Li,G.  
TITLE CDNA's from bovine abomasum tissue  
JOURNAL Unpublished (2001)  
COMMENT Contact: Dr. Stephen Moore  
. Beef Genomics Laboratory  
Dept of AFNS, University of Alberta  
410 Agri/For, Dept of AFNS, U of A, Edmonton, AB, T6G 2P5, Canada  
Tel: 780 492 0169  
Fax: 780 492 4265  
Email: smoores@afns.ualberta.ca  
The sequence best matches gb:BPJ19766 (Bos primigenius prochlorosin  
mRNA, complete cds) in main database at high score of 735.0 and  
E-value of 0.0  
PCR PRIMERS  
FORWARD: M13 Forward  
BACKWARD: M13 Reverse  
Seq primer: T3 primer  
High quality sequence stop: 383  
POLYA-NO.

#### FEATURES

source location/Qualifiers  
1..383  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/clone\_lib="Bovine Abomasum cDNA Library"  
/sex="Two males and one female mixed"  
/tissue\_type="Gastrointestinal tissue (GIT)"

Query Match	1.2%;	Score 49;	DB 10;	Length 383;
Best Local Similarity	100.0%;	Pred. No. 2.7e-08;		
Matches 49;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Oy	2209	tgcttcggtttacatgacagcaagatggccaggaagacatgctcacgct	2257	
Db	58	tgcttcggtttacatgacagcaagatggccaggaagacatgctcacgct	106	
RESULT	2			
LOCUS	BC937697	430 bp	mRNA	linear
DEFINITION	1AB005E01 Bovine Abomasum cDNA Library Bos taurus cDNA 5', mRNA			
ACCESSION	BC937697			
VERSION	BC937697.1	GI:14337069		
KEYWORDS	EST.			
SOURCE	cow.			
ORGANISM	Bos taurus			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;			
	Bovidae; Bovinae; Bos.			
	1 (bases 1 to 430)			
REFERENCE	Moore,S.S., Hansen,C., Li,C., Fu,A., Meng,Y. and Li,G.			
AUTHORS	CDNA's from bovine abomasum tissue			
TITLE	Unpublished (2001)			
JOURNAL	Contact: Dr. Stephen Moore			
COMMENT	Beef Genomics Laboratory			
	Dept of AFNS, University of Alberta			
	410 Agri/For, Dept of AFNS, U of A, Edmonton, AB, T6G 2P5, Canada			
	Tel: 780 492 0169			
	Fax: 780 492 4265			
	Email: smoores@afns.ualberta.ca			
	The sequence best matches gb:BOVGHYMOA (bovine chymosin a (rem1n)			
	min)in main database at high score of 844.0 and E-value of 0.0			
	PCR Primers			
	FORWARD: M13 Forward			
	BACKWARD: M13 Reverse			
	Seq primer: T3 primer			
	High quality sequence stop: 430			
	POLYA-No.			
FEATURES				
source	Location/Qualifiers			
	1..430			
	/organism="Bos taurus"			
	/db_xref="taxon:9913"			
	/clone_lib="Bovine Abomasum cDNA library"			
	/sex="Two males and one female mixed"			
	/tissue.type="gastrointestinal tissue (GIT)"			
	/cell_type="Epithelial"			
	/dev_stage="Young adult"			
	/lab_host="XLI-BlueMRP"-strain"			
	/note="Organ: Abomasum; Vector: Uni-22APXR; Site_1: Ecor			
	I; Site_2: Xho I"			
BASE COUNT	97 a 130 c 119 g 84 t			
ORIGIN				
Query Match	1.2%;	Score 49;	DB 10;	Length 430;
Best Local Similarity	100.0%;	Pred. No. 2.5e-08;		
Matches 49;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Oy	2209	tgcttcggtttacatgacagcaagatggccaggaagacatgctcacgct	2257	
Db	52	tgcttcggtttacatgacagcaagatggccaggaagacatgctcacgct	100	

[illegible]



```

RESULT 7
AG091550/c 660 bp DNA linear GSS 03-NOV-2001
Pan troglodytes DNA, clone: PTB-091H03.R, genomic survey sequence.
LOCUS
DEFINITION
ACCESSION AG091550
VERSION AG091550.1 GI:16643352
KEYWORDS GSS: GSS (genome survey sequence).
SOURCE Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
BAC library clone:PTB-091H03.R.
ORGANISM
Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.
REFERENCE
AUTHORS 1 (sites)
Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totohi,Y., Watanabe,H. and Sakaki,Y.
JOURNAL
TITLE BAC end sequences of library PTB
Unpublished
2 (bases 1 to 660)
Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totohi,Y., Watanabe,H. and Sakaki,Y.
REFERENCE
AUTHORS Direct Submission
Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou,Tsuriuni-Ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:chimbess@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT Clones are derived from the chimpanzee BAC library PTB This BAC end
was generated during the R&D process and may have higher chance of
clone tracking errors.
PRIMERS
Sequencing: M13Rev
LIBRARY
Vector : pKS145
R.Site 1 : SacI
R.Site 2 : SacI
Location/Qualifiers
1. 660
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="PTB-091H03.R"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC library"
BASE COUNT 128 a 166 c 130 g 236 t
ORIGIN
Query Match 0.6%; Score 25; DB 12; Length 660;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3403 ttaataataataataataata 3427
|||||
Db 94 TTAATAATAATAATAATAATAATA 70

```

```

RESULT 8
BE763309 136 bp mRNA linear EST 19-SEP-2000
LOCUS
DEFINITION RC1:NT0033-090800-016-a01 NT0033 Homo sapiens CDNA, mRNA sequence.
ACCESSION BE763309
VERSION BE763309.1 GI:10193233
KEYWORDS EST.
SOURCE human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS 1 (bases 1 to 136)
Dias Neto,E., Garcia Correa,R., Verjowski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,M. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baie,G.S., Simpson,D.H.,
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare

```

---

```

TITLE
JOURNAL
MEDLINE
COMMENT
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=ctc=RC1-NT0033-090
800-016-a01ct3-2000-08-09&tl=1)
Seq primer: puc 18 forward
High quality sequence start: 7
High quality sequence stop: 136.
Location/Qualifiers
1. 136
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NT0033"
/dev_stage="Adult"
/note="Organ: nervous-tumor; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 62 a 21 c 22 g 31 t
ORIGIN
Query Match 0.6%; Score 24; DB 10; Length 136;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3407 aataataataataataata 3430
|||||
Db 41 AATAATAATAATAATAATAATAACA 64

```

---

```

RESULT 9
C25503 214 bp mRNA linear EST 28-APR-1999
LOCUS
DEFINITION C25503 Dictyostelium discoideum SL (H.Urushi-hara) Dictyostelium
discoideum cDNA clone SLA138, mRNA sequence.
ACCESSION C25503
VERSION C25503.1 GI:2276053
KEYWORDS EST.
SOURCE Dictyostelium discoideum.
ORGANISM
Dictyostelium discoideum
Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.
REFERENCE
AUTHORS 1 (bases 1 to 214)
Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,
Yoshino,R., Mita,B.N., Pi,M., Sato,T., Takemoto,K., Yasukawa,H.,
Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.
The Dictyostelium developmental cDNA project: generation and
analysis of expressed sequence tags from the first-finger stage of
development
DNA Res. 5 (6), 335-340 (1998)
99156227
Contact: Hideko Urushihara
Institute of Biological Sciences
University of Tsukuba
3-3-10 Ten-nodai, Tsukuba, Ibaraki 305, Japan
Email: d402h@sakura.cc.tsukuba.ac.jp
PROJECT = 'Dictyostelium discoideum cDNA project in Japan'

```

FEATURES POLYA-NO. Location/Qualifiers  
SOURCE 1. .214  
/organism="Dictyostelium discoideum"  
/strain="AX4"  
/db\_xref="taxon:44689"  
/clone="SLJ138"  
/clone\_lib="Dictyostelium discoideum SL (H.Urushihara)"  
/dev\_stage="slug"  
BASE COUNT 82 a 32 c 28 g 72 t  
ORIGIN

Query Match 0.6%; Score 24; DB 10; Length 214;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3405 aaataataataataataataa 3428  
|||||  
Db 160 AAATAATAATAATAATAATAA 183

RESULT 10  
AU053870 312 bp mRNA linear EST 28-APR-1999  
LOCUS AU053870 Dictyostelium discoideum SL (H.Urushihara) Dictyostelium  
DEFINITION AU053870 Dictyostelium discoideum SL (H.Urushihara) Dictyostelium  
ACCESSION AU053870  
VERSION AU053870.1 GI:4702352  
KEYWORDS EST.  
SOURCE Dictyostelium discoideum.  
ORGANISM Dictyostelium discoideum  
REFERENCE Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.  
1 (bases 1 to 312)  
AUTHORS Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,  
Yoshino,R., Mitra,B.N., Pl.M., Sato,T., Takemoto,K., Yasukawa,H.,  
Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.  
Developmental cDNA in Dictyostelium discoideum  
Unpublished (1998)  
TITLE Contact: Hideko Urushihara  
JOURNAL Institute of Biological Sciences  
COMMENT University of Tsukuba  
3-3-10 Ten-nodai, Tsukuba, Ibaraki 305, Japan  
Email: d402nuesakura.cc.tsukuba.ac.jp  
PROJECT = Dictyostelium discoideum cDNA project in Japan.  
Location/Qualifiers  
1. .312  
/organism="Dictyostelium discoideum"  
/strain="AX4"  
/db\_xref="taxon:44689"  
/clone="SLJ838"  
/clone\_lib="Dictyostelium discoideum SL (H.Urushihara)"  
/dev\_stage="slug"  
BASE COUNT 110 a 50 c 50 g 102 t  
ORIGIN

Query Match 0.6%; Score 24; DB 9; Length 312;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3405 aaataataataataataataa 3428  
|||||  
Db 264 AAATAATAATAATAATAATAA 287

RESULT 11  
AA441444 442 bp mRNA linear EST 19-APR-2001  
LOCUS AA441444  
DEFINITION LD16384, Sprime LD Drosophila melanogaster embryo Bluescript  
Drosophila melanogaster cDNA clone LD16384 Sprime, mRNA sequence.  
ACCESSION AA441444  
VERSION AA441444.1 GI:4299185

KEYWORDS EST.  
SOURCE fruit fly.  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE 1 (bases 1 to 442)  
AUTHORS Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,  
Lewis,S. and Rubin,G.M.  
BDGP/HHMT Drosophila EST Project  
Unpublished (2001)  
JOURNAL On Jun 2, 1997 this sequence version replaced gi:2153322.  
COMMENT Contact: Stapleton, M.  
BDGP  
Lawrence Berkeley National Lab  
One Cyclotron Rd, Berkeley, CA 94720, USA  
Fax: 510 486 6798  
Email: <http://www.fruitfly.org/EST/estfruitfly.berkeley.edu>  
Plate: 163 row: G column: 12  
High quality sequence stop: 254.  
Location/Qualifiers  
1. .442  
/organism="Drosophila melanogaster"  
/db\_xref="BDGP-EST:BDc1n015599"  
/db\_xref="taxon:7227"  
/clone="LD16384"  
/clone\_lib="LD Drosophila melanogaster embryo Bluescript"  
/sex="male and female"  
/dev\_stage="0 to 24 hours mixed stage embryonic"  
/lab\_host="SOLR"  
/note="Organ: embryo; Vector: Bluescript SK; Site: 1: SCORI  
; Site: 2: XhoI; Constructed using Stratagene ZAP-cDNA  
Synthesis kit. Oligo dt-primed and directionally cloned at  
EcoRI and XhoI in Bluescript SK(+/-)"  
BASE COUNT 176 a 89 c 94 g 83 t  
ORIGIN

Query Match 0.6%; Score 24; DB 9; Length 442;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3404 taaataataataataataa 3427  
|||||  
Db 165 TAAATAATAATAATAATAATA 188

RESULT 12  
AU052887 500 bp mRNA linear EST 28-APR-1999  
LOCUS AU052887 Dictyostelium discoideum SL (H.Urushihara) Dictyostelium  
DEFINITION AU052887 Dictyostelium discoideum SL (H.Urushihara) Dictyostelium  
ACCESSION AU052887  
VERSION AU052887.1 GI:4701370  
KEYWORDS EST.  
SOURCE Dictyostelium discoideum.  
ORGANISM Dictyostelium discoideum  
REFERENCE Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.  
1 (bases 1 to 500)  
AUTHORS Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,  
Yoshino,R., Mitra,B.N., Pl.M., Sato,T., Takemoto,K., Yasukawa,H.,  
Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.  
Developmental cDNA in Dictyostelium discoideum  
Unpublished (1998)  
TITLE Contact: Hideko Urushihara  
JOURNAL Institute of Biological Sciences  
COMMENT University of Tsukuba  
3-3-10 Ten-nodai, Tsukuba, Ibaraki 305, Japan  
Email: d402nuesakura.cc.tsukuba.ac.jp  
PROJECT = Dictyostelium discoideum cDNA project in Japan.  
Location/Qualifiers  
1. .500  
/organism="Dictyostelium discoideum"

/strain="Ax4"  
/db\_xref="taxon:44689"  
/clone="SLP272"  
/clone\_1lb="Dictyostelium discoideum SL (H.Urushihara)"  
/dev\_stage="slug"  
BASE COUNT 169 a 88 c 83 g 160 t  
ORIGIN

Query Match 0.6%; Score 24; DB 9; Length 500;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3405 aaaaataataataataa 3428  
|||||  
DB 449 AAAATATATATATATATATA 472

RESULT 13  
BH366372 560 bp DNA linear GSS 03-DEC-2001  
LOCUS CH230-204G22.T1 CHORI-230 Segment 1 Rattus norvegicus genomic clone  
DEFINITION BH366372  
ACCESSION BH366372  
VERSION BH366372.1 GI:17297106  
KEYWORDS GSS.  
SOURCE Norway rat.  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
1 (bases 1 to 560)  
Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shvartsbeyn  
A., Gebregorjis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de  
Jong,P. and Frisner,C.M.  
Rat BAC End Sequences from Library CHORI-230 EcoRI segment  
Unpublished (1999)  
Other\_GSSs: CH230-204G22.TV  
Contact: Shaying Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: szhao@tigr.org  
Clones are derived from the rat BAC library CHORI-230  
(http://www.chori.org/bacpac/rat230.htm). For BAC library  
availability, please contact Pieter de Jong (pjejong@mail.cho.org).  
Clones may be purchased from BACPAC Resources  
(http://www.chori.org/bacpac/or\_eirlng\_information.htm). BAC end  
page: http://www.tigr.org/tldb/bac\_ends/rat/bac\_end\_intro.html  
Plate: 204 row: G column: 22  
Seq primer: SP6  
Class: BAC ends.

FEATURES  
source Location/Qualifiers  
1..560  
/organism="Rattus norvegicus"  
/strain="BN/SSNHsd/MCM"  
/db\_xref="taxon:10116"  
/clone="CH230-204G22"  
/clone\_1lb="CHORI-230 Segment 1"  
/sex="Female"  
/cell\_type="Brain"  
/note="Vector: pTARBAC2.1; Site.1: EcoRI; Site.2: EcoRI;  
CHORI-230 Rat (BN/SSNHsd/MCM) BAC library produced by  
Pieter de Jong"

BASE COUNT 195 a 103 c 89 g 173 t  
ORIGIN

Query Match 0.6%; Score 24; DB 12; Length 560;  
Best Local Similarity 100.0%; Pred. No. 98;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 398 catttataatttaaaatacatt 421  
|||||  
DB 311 CATTATTATTTTAAAAATATAT 334

RESULT 14  
AO635511 667 bp DNA linear GSS 17-JUN-1999  
LOCUS RPCI-11-490J20.TV RPCI-11 Homo sapiens genomic clone RPCI-11-490J20  
DEFINITION AO635511  
ACCESSION AO635511  
VERSION AO635511.1 GI:5098146  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 667)  
Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter  
J.C.  
Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready  
Map Building  
Unpublished (1997)  
Contact: Shaying Zhao, William Nierman, Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: hbeet@tigr.org  
Clones are derived from the human BAC library RPCI-11. For BAC  
library availability, please contact Pieter de Jong  
(pieter@dejong.med.buffalo.edu). Clones may be purchased from  
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from  
Research Genet cs (info@resgen.com). BAC end search page:  
http://www.tigr.org/tldb/humgen/bac\_end\_search/bac\_end\_search.html.  
Seq primer: T7  
Class: BAC ends.

FEATURES  
source Location/Qualifiers  
1..667  
/organism="Homo sapiens"  
/db\_xref="GDB:7688011"  
/db\_xref="taxon:9606"  
/clone="RPCI-11-490J20"  
/clone\_1lb="RPCI-11"  
/sex="Male"  
/cell\_type="Lymphocytes"  
/note="Vector: pBAC3.6; Site.1: EcoRI; Site.2: EcoRI;  
RPCI11 Human Male BAC library"

BASE COUNT 215 a 135 c 146 g 171 t  
ORIGIN

Query Match 0.6%; Score 24; DB 12; Length 667;  
Best Local Similarity 100.0%; Pred. No. 88;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3405 aaaaataataataataa 3428  
|||||  
DB 288 AAAATATATATATATATATA 265

RESULT 15  
BG872501 702 bp mRNA linear EST 29-MAY-2001  
LOCUS BG872501 NCL\_CGAP\_SG2 Mus musculus cDNA IMAGE:4924784 5',  
DEFINITION BG872501  
ACCESSION BG872501  
VERSION BG872501.1 GI:14223041  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 702)  
 AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 TITLE NIH-MGC <http://mgc.nci.nih.gov/>.  
 JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)  
 COMMENT Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
 Tissue Procurement: Jeffrey E. Green, M.D.  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.llnl.gov>  
 Plate: LLM10848 row: e column: 09  
 High quality sequence stop: 655.  
 Location/Qualifiers

## FEATURES

## Source

1..702  
 /organism="Mus musculus"  
 /strain="FVB/N"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:4924784"  
 /clone\_lib="NCI\_CGAP\_SG2"  
 /lab\_host="DH10B (T1 phage-resistant)"  
 /note="Organ: salivary gland; Vector: PCMV-SPORT6; Site 1:  
 NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo  
 dT. Average insert size 1.3 kb. Constructed by Life  
 Technologies. Note: this is a NCI\_CGAP Library."  
 BASE COUNT 220 a 149 c 152 g 181 t  
 ORIGIN

Query Match 0.6%; Score 24; DB 10; Length 702;  
 Best Local Similarity 100.0%; Pred. No. 85;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1109 aaagacaagaagaaagacaaga 1132  
 ||||||||||||||||||  
 Db 666 AAAGACAAGAAAAGACAAGA 689

Search completed: July 31, 2002, 23:07:38  
 Job time: 17283 sec

**THIS PAGE BLANK (USPTO)**